Estadística aplicada

Alejandro Urieles Guerrero
William Ramírez Quiroga
Julio Cesar Romero Pabón

libro digital

CUC
CORPORACIÓN
UNIVERSIDAD
DE LA COSTA
1970
VIGILADA MINEDUCACIÓN

Estadística aplicada

Romero Pabón, Julio César
Estadística aplicada /Julio César Romero Pabón, Alejandro Urieles Guerrero, William Ramírez Quiroga. – Barranquilla, 2017
232 páginas
ISBN: 978-958-8921-59-4 (Digital)
1. Estadística 2. Variables (Estadística)
3. Muestreo (Estadística)

310 R763 Co-BrCuC

Estadística aplicada

Alejandro Urieles Guerrero
William Ramírez Quiroga
Julio Cesar Romero Pabón

CORPORACIÓN UNIVERSIDAD DE LA COSTA 1970
VIGILADA MINEDUCACIÓN

2017

EDUCOSTA
EDITORIAL CORPORACIÓN UNIVERSIDAD DE LA COSTA

Estadística aplicada

Autor: Alejandro Urieles Guerrero
William Ramírez Quiroga
Julio Cesar Romero Pabón

CORPORACIÓN UNIVERSIDAD DE LA COSTA
Barranquilla - Colombia - Sur América

ISBN: 978-958-8921-59-4 (Digital)

Primera Edición
Editorial Corporación Universidad de la Costa,
EDUCOSTA
Departamento de Gestión Editorial y Publicaciones
Corporación Universidad de la Costa
Calle 58 No. 55-66
Teléfono: (575) 336 2272
educosta@cuc.edu.co

Lauren J. Castro Bolaño
Directora Departamento de
Gestión Editorial y Publicaciones

Carolina Mercado Porras
Auxiliar Departamento de
Gestión Editorial y Publicaciones

Corrección de Estilo
Diagramación
Diseño de Portada:
Dolores López

Hecho el depósito que exige la ley.

FUNDADORES

EDUARDO CRISSIÉN SAMPER
RUBÉN MAURY PERTUZ (q.e.p.d)
NULVIA BORRERO HERRERA
MARÍA ARDILA DE MAURY
RAMIRO MORENO NORIEGA
RODRIGO NIEBLES DE LA CRUZ (q.e.p.d)
MIGUEL ANTEQUERA STAND

PERSONAL DIRECTIVO

TITO JOSÉ CRISSIÉN BORRERO
Rector

GLORIA CECILIA MORENO GÓMEZ
Vicerrectora Académica

HENRY MAURY ARDILA
Vicerrector de Investigaciones

JORGE MORENO GÓMEZ
Vicerrector de Extensión

JAIME DÍAZ ARENAS
Vicerrector Administrativo

ROSMERY TURBAY MIRANDA
Vicerrectora de Bienestar

HERNANDO ANTEQUERA MANOTAS
Vicerrector Financiero

ALFREDO GÓMEZ VILLANUEVA
Decano Facultad
de Arquitectura

JAVIER MORENO JUVINAO
Decano Facultad de
Ciencias Económicas

JOSÉ LOZANO JIMENEZ
Decano Facultad de
Ciencias Sociales y Humanas

ALFREDO PEÑA SALOM
Decano Facultad de Derecho

FAIRUZ OSPINO VALDIRIS
Decano Facultad de Ingeniería

CARMEN MEZA ESTRADA
Directora Departamento
Arquitectura y Diseño

LIGIA ROMERO MARÍN
Directora Departamento de
Derecho y Ciencias Políticas

NOEL VARELA IZQUIERDO
Director Departamento
Gestión Industrial,
Agroindustrial y Operaciones

LISETTE HERNÁNDEZ FERNÁNDEZ
Directora Departamento de
Gestión Organizacional

ALICIA INCIARTE GONZÁLEZ
Directora Departamento de
Humanidades

MARÍA DEL MAR SÁNCHEZ
Directora Departamento de
Psicología del Individuo

MARINA MARTINEZ GONZÁLEZ
Directora Departamento de Psicología
de las Interacciones Sociales

JENNY ROMERO DE CUBA
Directora Departamento Economía,
Contabilidad y Finanzas

JUAN CABELLO ERAS
Director Departamento de Energía

EMIRO DE LA HOZ FRANCO
Director Departamento de Ciencias
de la Computación y Electrónica

ALDEMAR DE MOYA CAMACHO
Director Departamento de
Ciencias Naturales y Exactas

LUIS SILVA OLIVEIRA
Director Departamento de
Civil y Ambiental

DEDICATORIA

A nuestros padres:
Con respeto y admiración a nuestros padres; por su fe, cariño, sacrificio, esfuerzo, enseñanza, orientación y ejemplo. Porque han infundido en nosotros sabiduría de las sagradas escrituras, haciendo que hoy tengamos el conocimiento de lo que somos.

A nuestros hermanos:
Agradezco a mis hermanos el apoyo que siempre me han brindado con su impulso, fuerza y tenacidad que son parte de mi formación. Como muestra de gratitud les dedico éste trabajo.

A nuestras esposas:
Por su amor, por todo su apoyo y confianza.

A nuestros hijos:
Porque son el aliento que nos da ánimo para seguir siempre adelante.

AGRADECIMIENTOS

A Dios:
Por su misericordia y protección, por las bendiciones recibidas, por ayudarme a terminar un proyecto más en mi vida, y por darme las fuerzas para seguir adelante. Gracias Dios mío por estar siempre con nosotros.

A nuestras familias:
Por su amor y apoyo recibido.

A los profesores y directivos de la Universidad del Atlántico y de la CUC:
Por su amabilidad, confianza, cooperación y todas las atenciones que han tenido con nosotros; su ayuda fue de gran utilidad en el desarrollo de éste trabajo.
Con todo respeto y gratitud a quienes de una u otra manera me han apoyado y me han transmitido su valioso conocimiento y experiencia, también por todas las facilidades prestadas para la realización de éste proyecto.

A nuestros amigos:
Por compartir sus experiencias en el transcurso de este proyecto de investigación. Por su apoyo incondicional.

A la Universidad del Atlántico y a la CUC:
Por brindarnos la oportunidad de ampliar y difundir nuestros conocimientos en tan prestigiosas instituciones.

CONTENIDO

CAPITULO III

CAPITULO IV

CAPITULO V

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

PRÓLOGO

El propósito fundamental de este libro es presentar la estadística desde el punto de vista de sus aplicaciones, por medio de un modelo didáctico que propende que el estudiante o lector comprenda bien cada uno de los tópicos expuestos en esta obra académica. Es bien conocido por todos que la estadística es una herramienta fundamental en los procesos de investigación e informes que se presentan ante las empresas o instituciones con el objeto de apreciar, describir e inferir sobre un evento o caso de estudio.

En esta era del conocimiento y de la información no se puede concebir una investigación en las ciencias experimentales, en la medicina, en las ciencias sociales, en las técnicas o en la industria, que no utilice la estadística para analizar la información y así poder ordenar, clasificar, describir e inferir acerca de lo que se está investigando. Es por eso que el estudiante y el investigador de hoy necesitan de los métodos estadísticos para poder aplicarlos en su labor o proyecto.

En este libro se tratan todos los temas de la estadística descriptiva y algunos tópicos fundamentales de la probabilidad, como son las técnicas de conteo, sus aplicaciones y el origen y desarrollo de las probabilidades según sea el caso. Todos estos temas son explicados con conceptos, análisis y muchos problemas de aplicaciones prácticas.

De antemano reconocemos y agradecemos el esfuerzo de todas aquellas personas que han contribuido a que esta obra se haya podido terminar; a nuestros profesores que nos han ayudado a formarnos, a los compañeros, principalmente a los de la facultad de ciencias básicas por apoyarnos en la construcción de este libro de estadística, el cual será una herramienta importante para los proyectos presentados por estudiantes, docentes e investigadores.

Los Autores

PRESENTACIÓN

Esta obra está orientada a fortalecer los procesos de enseñanza y aprendizaje de la estadística en las instituciones de educación superior; proporcionando a los estudiantes y docentes de un material bibliográfico para sus investigaciones o trabajos. Este libro contempla métodos y técnicas útiles para todos los campos del conocimiento, porque con ellos es posible comprender y aplicar tópicos estadísticos como: recopilar, planificar, organizar describir y realizar inferencias de una investigación o estudio.

La intención de esta obra es la de presentar un material de estadística básica para los interesados en comprender los principios estadísticos que tanto se usan y se observan en la realidad, y a los estudiantes que esten interesados en la estadística como disciplina y un instrumento útil en otras áreas de conocimiento.

El mayor esfuerzo de este proyecto sobre la enseñanza y aprendizaje de la estadística va dirigido al estudiante, pues en el libro los temas son presentados con una gran variedad de estrategias encaminadas a brindar un escenario significativo para el aprendizaje de la estadística. Es por estas razones que los ejemplos fueron diseñados y seleccionados para que el estudiante o lector los comprenda y aplique fácilmente en cada uno de los tópicos tratados en esta obra.

Los profesores se encargaron de revisar la teoría, los ejemplos, los talleres y aplicaciones, para así contribuir con este proyecto de estadística. Es por esto que hoy consideramos conveniente presentar esta obra, porque cada uno de sus capítulos fue elaborado para que el lector conciba y aplique los temas de estadística tratados.

INTRODUCCIÓN

En el mundo actual las cifras, evoluciones de los sistemas, variaciones en los índice de precio, gastos familiares, índices económicos, bolsa de valores, censo electoral, porcentaje de personas que padecen una enfermedad, número de accidentes de circulación, tasa de natalidad y mortalidad, eficacia en publicidad, porcentaje de hogares según el estrato económico, prospecciones petrolíferas y minas, gastos en los servicios públicos y privados, informes administrativos y de gestión, son algunos ejemplos de la importancia que tiene la estadística en nuestra era.

Para Hernández (2013) proviene del término alemán *statistik*, que fue introducido por Gottfried Achenwall en 1749, para referirse al análisis de datos del Estado, es decir, la *ciencia del Estado* (o más bien, de la ciudad-estado). También conocida como la aritmética política de acuerdo con la traducción literal del inglés. No fue hasta el siglo XIX cuando el término adquirió el significado de recolectar y clasificar datos, al ser introducido este concepto en las obras del militar británico Sir John Sinclair (1754-1835). En el Imperio Romano se utilizaba para enumeración y recuento de soldados, medios de transporte, riquezas, etc. Además, se tiene constancia que en tiempos de César Augusto se realizó un censo de la población (Estrella, 2011). Otro origen según Hald (2003) deriva de la palabra *De status*, o estado de las cosas, la que para muchos ha sido considerada el origen de la estadística descriptiva.

Históricamente la estadística se ha considerado como una rama de las matemáticas, la cual ha tenido un origen remoto y en continua evolución y desarrollo. El término alemán Statistik, fue introducido por Gottfried Achenwall en 1749, para referirse al análisis de datos del Estado, es decir, la «ciencia del Estado» (o más bien, de la ciudad-estado). También fue conocida como la aritmética para la organización política. Otras civilizaciones como el Imperio Romano aportaron el sistema de enumeración y recuento de personas o soldados, medios de comunicación y transporte, riquezas, etc. Además se tiene evidencia que en el reinado de César Augusto se ejecutó un censo para analizar la población, de donde se considera derivarse la palabra estadística, o lo que muchos han considerado el origen de la estadística descriptiva. (Hald, 2003).

La estadística se ocupa de la construcción de modelos a partir de los fenómenos que evolucionan de forma aleatoria y se convierte en una manera de ayudar a la toma de decisiones que antes estaban sujetas a la incertidumbre, al ser capaz de simbolizar mediante esquemas las probablidades, convirtiéndose en una herramienta indispensable para la cimentación de la ciencia. Al ser los métodos estadísticos objetivos y útiles para todos los campos del conocimiento, podemos recopilar, organizar, describir y realizar inferencias en los estudios, volviendo indispensable su enseñanza en las instituciones de educación superior.

CAPITULO I

CONCEPTOS GENERALES

INTRODUCCIÓN

Algo muy particular en casi todas las personas, es que cuando inician cualquier tipo de actividad o negocio, se ven obligados a tomar o calcular ciertos valores o medidas, las cuales consideramos muy importante para la administración de estas actividades. Esta intuición o idea primaria es el motor de progreso de cualquier negocio o actividad que desarrollemos en vida diaria. Comprender y manejar la dinámica de estos sucesos, nos permitirá estudiar y analizar las similitudes, diferencias y regularidad de un modelo estadístico que rige el sistema donde estamos trabajando.

Este análisis ha hecho que el hombre vea necesario contar, calcular y evaluar estos sucesos que nos rodean, y al hacerlo es inevitable el hecho de "cuantificar", y con base en estas observaciones y comparaciones, se pueden resolver interrogantes como son:, cuánto pasa?, con qué frecuencia?, qué tan rápido?, qué tan bien se hace?, qué tan grande o pequeño?, qué tan cerca o lejos?, etc., sobre los hechos o sucesos que analizamos o experimentamos.

Así por nuestras experiencias podemos hacer una o varias observaciones de un objeto, individuos, o de varios individuos, y una a una, se obtienen una serie de observaciones, conocidas como *datos estadísticos*, y con características comunes, conexiones con otros datos, similitudes, etc., relacionadas entre sí y utilizadas en la medida que permiten dar respuestas a las preguntas mencionadas antes.

Esta manera de conectar las características similares, de distinguir referencias significativas, de aprender de las experiencias cotidianas, y que en una u otra forma ayudan a actuar más eficaz y efectivamente en el futuro, es cuando empieza a tener aplicación y utilidad el concepto de manipulación y tratamiento de datos, que es en esencia de lo que trata el estudio de la estadística, que a través de estas notas se inicia, y que nos permite ponerle sentido a todas estas colecciones de observaciones y características de fenómenos naturales, que a diario contemplamos.

1. RAMAS

En realidad, es sumamente difícil lograr una apreciación general sobre el concepto de estadística. Sin embargo, todos los estadísticos están de acuerdo en clasificar esta ciencia en dos grande ramas o partes: la estadística descriptiva y la estadística inferencial, las cuales son en realidad las dos funciones principales del análisis estadístico.

Analizaremos cada una de estas dos ramas del conocimiento estadístico, y, trataremos de justificar y dar las razones pertinentes para su estudio y de las múltiples aplicaciones que de su conocimiento se derivan.

a) Descriptiva

Cuando escuchamos la palabra "estadística" rápidamente pensamos en promedio de datos, índice o número de accidentes, tasas de mortalidad y nacimiento. En realidad, la rama de la estadística descriptiva utiliza números para describir, organizar y presentar la información que se recolecto de los hechos investigados.

Fundamentalmente, la estadística descriptiva se encarga principalmente de: el diseño y aplicación de instrumentos de investigación, su recolección, ordenamiento, conteo o tabulación, tratamiento matemático de los datos, análisis de los resultados con el objeto de presentar un buen informe que ilustre las situaciones que han arrojado la información recolectada. Esto implica por lo general la construcción de tablas, gráficos, tasas o índices numéricos, proporciones, etc. Todos estos factores permiten describir detalladamente lo que está sucediendo a una actividad o una investigación realizada en cualquier campo del saber.

Su objetivo es dar una información más detallada, con características descriptivas, de fácil comprensión, interpretación y análisis. Este análisis sólo se limita a los datos obtenidos de las observaciones hechas entre ellos, sin intervenir en nada a generalizaciones o inferencia acerca de todas las observaciones posibles.

Así, por ejemplo, si el departamento de personal de una empresa hace una investigación sobre los ingresos de sus empleados, entre las muchas cosas que podría hacer la oficina de personal con los resultados obtenidos estarían los siguientes aspectos:

- Ordenar y clasificar los salarios (de mayor a menor) de tal manera que los datos puedan dar un concepto general sobre la composición salariar de las empresas.

- Calcular el promedio de los ingresos, y su representatividad con el total de los empleados.

- Construir tablas, cuadros y gráficos con una visión de los niveles salariales de los empleados.

- Convertir los salarios (individuales) a salarios por categorías o rangos y calcular percentiles con el objeto de hacer comparaciones salariales.

- Utilizar el promedio como punto central de referencia para describir la dispersión o variación de los ingresos entre empleados.

- Si es posible, medir las relaciones existentes entre el rendimiento en el trabajo y su nivel salarial con base en los datos obtenidos sobre el rendimiento salarial.

- En la medida que se puede establecer una relación entre el rendimiento laboral y su nivel de ingreso, se puede predecir (en cierta medida) el salario que debería devengar una persona, basado en su rendimiento en el trabajo.

- Y otros aspectos, que permitan describir en una forma objetiva y amplia lo referente a la actividad laboral.

b) Inferencial

Una descripción pormenorizada de los datos recolectados de una actividad en particular es a veces el objetivo fundamental en una investigación; sin embargo, en la mayoría de los análisis estadísticos, el investigador se encuentra más al comienzo de su investigación que al término de la misma cuando ha hecho toda la descripción de los sucesos, y en realidad esto es así puesto que la finalidad de la labor estadística es sacar conclusiones útiles y valederas sobre la totalidad de las observaciones o el universo analizados. Esto es para la estadística descriptiva un trabajo preliminar para la inferencia estadística (Martínez, 2009).

Antes de proceder, a dar un concepto generalizado de lo que debe entenderse por inferencia estadística, analicemos dos conceptos básicos y fundamentales para este campo: *población y muestra.*

Estos conceptos son de mucha importancia en la estadística, ya que deben definirse muy cuidadosamente con el objeto de poder determinar la pertenencia que existe de ellos.

Población:

La *población* es considerada como una colección completa de individuos, objetos, medidas, que poseen una característica en común (Navidi, 2006). Así, por ejemplo, en el ejemplo anterior con los salarios de todos los empleados de la empresa, estos constituyen la población objeto de estudio y alguna situación diferente, u otras mediciones de variables producen una población distinta.

Las características principales de una población se suelen tomar generalmente como sus "parámetros". Un parámetro es un número que representa algo muy significativo en una población. Ejemplo, el salario promedio de los empleados de una fábrica, departamento o país, este promedio es concebido como una característica o parámetros de la población que se esté analizando (fábrica, departamento o país).

Pero, en la práctica, calcular este verdadero valor poblacional, con un número tan grande de datos es bastante impracticable. Comúnmente en todas las investigaciones estadísticas se utilizan la recolección y colección de información parcial a la cual denominan *muestra*, con base en la cual se hacen inferencias acerca del valor verdadero del parámetro de la población estudiada.

Muestra:

Se considera como un subconjunto de un universo o población, es decir un conjunto pequeño o colección personas, animales, plantas, objetos o cosas tomadas de una población. Las medidas o cantidades representativos que se obtengan de la muestra se denomina un "estadígrafo".

Ejemplo si se escogen 5000 personas de 10.000.000 de obreros de empresas públicas que tiene Colombia, y se les consulta sobre los ingresos anuales a estas 5000 personas que son la muestra, entonces el promedio de sus ingresos anuales será un estadígrafo

La *Inferencia Estadística* es una rama de la estadística que obtener conclusiones generalizadas acerca de un parámetro de una población. Para esto es necesario la selección de una muestra, de donde se obtendrán los estadígrafos que servirán como referentes para poder hacer inferencias en los parámetros de la población.

Sin embargo, existe una cierta posibilidad de que esta generalización pueda ser errada. No sería necesario ninguna inferencia estadística respecto del parámetro de una población si se han tomado todas las posibles observaciones de la población. Pero debido a las limitaciones de tiempo, de recursos económicos y de equipos, se hace forzoso decidir respecto al valor del parámetro de una población sobre la base del valor de estadígrafo de la muestra. Esto da una indicación del papel que desempeña la inferencia estadística en la investigación científica moderna.

Así, vemos que las empresas privadas y las dependencias gubernamentales realizan muestreo por múltiples razones, y el costo suele ser el factor principal. Como cualquier otra actividad, recopilar datos y analizar los resultados cuesta dinero y en cuanto se utiliza una gran cantidad de datos, mayor será el costo, por tanto, el uso de muestreo reduce la cantidad de datos que se deben recopilar y analizar, haciendo que el costo se reduzca sustancialmente.

Otra razón para utilizar el muestreo en vez del estudio de toda la población, es el hecho de que la información recolectada pierde valor y validez en muy poco tiempo, puesto que la población es muy activa y está constantemente cambiando. Para que la población sea útil, esta se debe obtener y aprovechar con rapidez y el muestreo cumple muy bien con estos objetivos.

Otra razón valedera del uso del muestreo es cuando se tienen que examinar o someter a prueba muchos artículos y estos se destruyen o deterioran perdiendo sus cualidades físicas, químicas, etc. Así, por ejemplo, al examinar una producción de bombillos con el objetivo de medir su durabilidad, evidentemente se funden algunos de los mismos. Si toda la producción se examinara mediante este proceso, no quedaría ninguna bombilla en buen estado para vender o probar. Por tanto, es de lógica utilizar una parte de la producción y se logran los mismos objetivos como si se analizara toda la producción.

2. DEFINICIÓN

Analizadas las consideraciones relacionadas con las dos principales ramas de la estadística (descriptiva e inferencial), esta se define en los siguientes términos:

La *Estadística* se concibe como una ciencia que hace uso de teorías, técnicas y métodos altamente especializados para la recolección, organización, tabulación, representación gráfica, descripción y análisis de la información seleccionada mediante muestreos, con el objeto de obtener buenas conclusiones que permitan su utilidad y valides para que sean aplicadas a la población de donde se extrajo la muestra. Todo esto conlleva cumplir con un alto grado de confiabilidad enmarcado dentro de un espacio y tiempo pertinente

Su función primordial es apoyar al investigador cuando esta toma decisiones y plantea estrategias sobre unos parámetros de la población.

En la realidad, la estadística es una herramienta bastante útil y eficaz en la toma de decisiones de aquellos estados naturales que se presentan bajo incertidumbre, lo cual se manifiesta en dos formas:

- *Modelos probabilísticos teóricos*, tales como el lanzamiento de una moneda, un dado, naipes, etc.
- *Hechos que no obedecen a una ley aleatoria*, como los fenómenos genéticos, los sociales, agrícolas, físicos, meteorológicos, etc.

3. METODOLOGÍA

Como ciencia que es la estadística utiliza el método científico, el cual podemos resumir en sus cincos pasos básicos indispensables en todo proceso de metodología estadística. Estos pasos básicos son los siguientes, en orden:

- *Definir bien y cuidadosamente el problema o investigación que se desea adelantar*. Se debera asegurar de que los objetivos que se persiguen están definidos con bastante claridad.
- *Formular un plan o planificación de las tareas a ejecutar* para la adecuada recolección de los datos pertinentes y necesarios.
- *Recolección, organización, tabulación de los datos estadísticos*.
- *Análizar e interpretar* los resultados obtenidos durante la investigación.
- *Sacar conclusiones útiles y valederas,* de fácil comprensión para las personas que empleen los resultados en la toma de decisiones o planeación de estrategias.

La investigación estadística a través de los métodos científicos se ocupa de sacar conclusiones de los experimentos o investigaciones que han sido cuidadosamente planeadas (Freund, Miller y Miller, 2000). Por lo tanto, las conclusiones van desde lo particular hasta lo general, a través de métodos estadísticos controlados, que llevan a conocer el comportamiento de los elementos de la población, con un alto grado de precisión y confiabilidad.

No debe perderse de vista el hecho de que la estadística sólo es una herramienta de ayuda a la investigación, de tal manera que no es por medio de ella que se va a razonar científicamente en busca de la solución del problema que se está tratando. Pero en gran parte ayuda mucho al investigador a planear y orientar sus problemas para que se facilite la apreciación objetiva de los resultados numéricos y poder avivar a la mejor conclusión de la mejor experimentación.

4. RAZONES PARA ESTUDIAR ESTADÍSTICA

Si existe una función que sea la más importante universalmente para aquella persona que labora dentro de una organización empresarial, es la *decisión.* Debido al enorme aumento de disponibilidad de datos a través del internet, y dada la gran complejidad del proceso de decisión se somete a presiones extraordinarias a los responsables de las decisiones y tienen que estar familiarizadas con las técnicas estadísticas existentes, para poder determinar cuándo se puede analizar una situación mediante la aplicación de la estadística.

La mayoría de las personas que toman decisiones efectúan realmente el análisis, ya que esta operación es del dominio del estadístico empresarial o especialista, pero el ejecutivo que decide en una empresa debe tener el nivel de conocimiento necesario para comprender tal análisis. La necesidad de tal conocimiento de estadística no se limita a la persona que decide en una empresa, sino que todos deben estar al tanto de las técnicas actuales de esta ciencia. No hay manera de escapar hoy en día a nuestra calidad de consumidores de análisis estadísticos.

Este es el objetivo primordial a través de las principales técnicas estadísticas en las siguientes páginas de este libro. Los estudiantes, al iniciar un curso de estadística se preguntan, ¿por qué debo aprender estadística? En estos tiempos, los estudiantes indagan por la utilidad práctica que les debe proporcionar todo curso nuevo que ellos se ven abocados en sus estudios superiores, trataremos de satisfacer convenientemente estas inquietudes.

5. ÁREAS DE APLICACIÓN

La estadística es una de las ciencias que más se aplican en la actualidad. A continuación, se relacionan algunos ejemplos donde utilizan la estadista como una herramienta fundamental.

- **Agricultura**

La estadística cuenta con ciertos modelos que permiten analizar experimentos como son: la reproducción de seres vivos, estudio y análisis de fertilizantes, insecticidas, métodos para aumentar el rendimiento de cultivos o poblaciones.

- **Biología**

Son utilizados varios métodos estadísticos que facilitan estudiar las reacciones e interacciones entre las plantas y los animales frente a diversas situaciones ambientales.

- **Negociaciones**

El hombre de negocio usa la estadística para predecir las ventas, estimar las tendencias de los consumidores ante los productos, tomar buenas decisiones para poder invertir en sus actividades comerciales que son determinantes para el futuro de su empresa.

- **Salud y medicina**

La investigación sobre la efectividad y efectos de un nuevo medicamento o fármacos requieren de diseño experimental. Los técnicos de la salud la utilizan la estadística para planear sus estrategias logísticas en los hospitales y en otras instituciones prestadoras de salud.

- **Industria**

Los industriales utilizan la estadista para el control de calidad de sus productos, los cuales requieren de técnicas estadísticas basadas en el análisis descriptivo e inferencial.

- **Psicología**

Los psicólogos usan la estadística para sus estudios o investigaciones, lo cual requiere un dominio de conceptos y técnicas estadísticas que permitan analizar y comprobar la conducta, actitudes, inteligencia y su personalidad.

- **Sociología**

En sociología la estadística permite realizar estudios de comparación de grupos clasificados por factores como son: sociales, socioeconómicos, culturales, educativo y de compartimiento o relacional.

6. OBJETIVOS DE LA ESTADÍSTICA

Al tratar con datos estadísticos siempre es necesario extraer la máxima información de dichos datos, aquí se enumeran algunas de las posibilidades que se presentan en esa magnitud:

- Calcular los promedios aritméticos o medidas. Estas medidas proporcionan una indicación del comportamiento característico del grupo.
- Determinar la variabilidad de las observaciones tomando la medida como referencia, haciendo posible determinar la dispersión de los datos u observaciones en dicho valor central.
- Elaborar gráficos, diagramas, tablas, cuadros, figuras, etc., para describir con claridad la naturaleza de distribución de los datos.
- Fijar la relación existente entre las variables. A estos parámetros se les denomina *coeficientes de correlación* y son de gran utilidad en toda investigación estadística.
- Definir la *fiabilidad* de los instrumentos de medición, por ejemplo, al hacer dos series de mediciones de una misma persona con un mismo instrumento, o con dos instrumentos semejantes y

luego hallar la correlación entre ambas series de datos, o medir cuál de las dos series es homogénea.

- Especificar la validez de las mediciones. La correlación entre puntuaciones obtenidas según un Test o prueba, y las obtenidas en otra medición, llamada *criterio*, es un índice de validez.
- Emplear una serie de mediciones, o combinación de variables, para predecir el compartimiento futuro.
- Formular deducciones acerca de la población de procedencia de las muestras, lo que se constituye como la *inferencia estadística*, la cual es una de las principales actividades de la investigación estadística moderna.
- Comparar la actuación entre grupos y comprobar el significado de cualquier diferencia entre los parámetros considerados en cuestiones entre ellos. Por ejemplo, comparar los promedios de dos grupos, al determinar sus diferencias de una manera significativa, y no es posible una inspección. Sin embargo, se puede hacer un contraste de significación para verificar si la diferencia se puede atribuir a una variación casual. Los procedimientos de inferencia, de predicción, y de contraste de significación son ejemplos de la teoría de muestra o de la inferencia estadística.

7. ABUSOS DE LA ESTADÍSTICA

Cualquiera que trabaje en alguna de las ramas de la estadística siempre debe estar alerta para evitar cualquier mal uso de la recopilación de datos. El empleo erróneo de los datos o un razonamiento descuidado e ilógico pueden desorientar al investigador y destruir el valor de un estudio.

Con frecuencia, se da el caso que las conclusiones se pueden basar en datos muéstrales que no son representativos así que las conclusiones que se saquen de esta mala utilización de los datos conducirá inexorablemente a un razonamiento desacertado en una investigación. Por consiguiente, hay que ser muy cuidadoso en la elección y perfeccionamiento de los procedimientos muéstrales empleados para poder garantizar el carácter representativo de las muestras.

También, con frecuencia se pueden llegar a conclusiones erróneas debido a que los datos recopilados son insuficientes en cifras. Así, por ejemplo, suponiendo que un vendedor de pólizas de seguros hace visitas de casa en casa para realizar sus ventas y despúes de analizar las cifras el 25 por ciento de sus visitas arrojan ventas efectivas. Pero su interpretación se basa en que efectua una sola venta un sábado por la tarde de un total de 4 visitas en ese día, lo cual da la impresión de que el 25 por ciento de todas sus visitas tienen éxito, y evidentemente no es así. Su pretensión tendría más validez si se diera el número total de visitas durante un período más largo de tiempo. Habría más inclinación a creerle si dijera que de 100 visitas a la semana, 25 dieron éxitos de venta, o aún mejor, si dijera que de 1000 visitas al mes 250 tuvieron éxito. Cuanto mayor sea el número abarcado más confiable será la proporción de la muestra. Así, pues, cuando se utilice una fracción, o un tanto por ciento, es más aconsejable conocer el número total de casos o de observaciones para poder garantizar un alto nivel de confiabilidad de los resultados.

AUTOEVALUACIÓN

1. Defina cada uno de los siguientes conceptos:

- Estadística
- Muestra y población
- Estadística descriptiva y sus rasgos distintivos
- Estadística inferencial y sus rasgos distintivos

2. ¿Cuáles son las ventajas de utilizar una muestra en un proyecto de investigación?
3. ¿Qué pasos se deben realizar para analizar estadísticamente los datos obtenidos en una investigación?
4. Diga 5 razones por las cuales es importante conocer y manejar la estadística.
5. De 5 ejemplos donde se aplica la estadística descriptiva y la inferencial.
6. ¿Qué diferencia hay entre la estadística descriptiva y la inferencial?
7. De 4 ejemplos de situaciones donde se habla de un abuso o mal uso de los datos estadísticos.

CAPITULO II

ORGANIZACIÓN Y PRESENTACIÓN DE DATOS ESTADÍSTICOS

INTRODUCCIÓN

Los modelos estadísticos expuestos en la unidad anterior comprenden primordialmente el análisis e interpretación de las cifras, a las que denominaremos *datos estadísticos*. Tales datos como las notas de un examen, las estaturas, peso de los estudiantes, números de partes defectuosas de un proceso industrial, años de servicio de una empresa, porcentaje de respuestas correctas a un cuestionario, ventas mensuales de un producto, etc., son de los muchos ejemplos que constituyen los datos estadísticos.

Por lo general, con el objeto de poder interpretar correctamente, y extraer de ellos una mejor y adecuada información, es necesario organizarlos y presentarlos en forma tabular o gráfica. Precisamente el objeto de esta unidad es demostrar los métodos más comúnmente utilizados para su organización y presentación, y que servirán de base a un posterior análisis de los resultados obtenidos.

1. TÉRMINOS COMUNES

Variable:

Es una característica o fenómeno que puede tomar diferentes valores. Ejemplo: peso, talla, calificaciones de álgebra, puesto en el grupo escolar.

Constante:

Al contrario de la variable una constante no tiene sino un único valor. Ejemplo: el valor de $\pi = 3{,}1416$.

Datos:

Son números o medidas que han sido recopilados como resultado de observaciones. Los datos pueden ser *cualitativos* si el valor asignado es un atributo, por ejemplo color, sabor, sexo, etc. y *cuantitativos* si el valor asignado es un número real, por ejemplo edad, peso o estatura. Ejemplo: El número de ciudadanos que prefieren un candidato a la presidencia, las notas de historia de los estudiantes de octavo, los puntajes de admisión a la Universidad.

Población:

Como ya dijo Navidi (2006) en el capitulo anterior, *población* es el conjunto completo de individuos, objetos o medidas que poseen una característica común observable. Ejemplo: todos los ciudadanos de un país en edad de votar. Los egresados de una Universidad.

Muestra:

Tambien Canavos (1999) explica en el capítulo anterior como un subconjunto de la población o universo se conoce con el nombre de *muestra*. Ejemplo: Los egresados de una universidad entre los años 1994 a 1996.

2. CLASIFICACIÓN DE LAS VARIABLES

Los datos no procesados, se obtienen en forma bruta, o de un listado general de características medibles en una encuesta, careciendo de utilidad y significado si no se procesan en forma ordenada y organizada. Estas grandes cantidades de números, tienden más bien a confundir en lugar de aclarar, ya que la mente por sí sola no puede mejorar la diversidad de detalles y características medibles que acompañan a este gran conjunto de datos.

El procesamiento estadístico transforma los datos en información, organizándolos y condensándolos en cuadros y gráficas, o en unos cuantos números indicados que revelan su esencia. Su propósito es eliminar los detalles menores y hacer resaltar las características importantes de los datos. Las *gráficas* no solamente sirven como instrumentos de comunicación, sino que también ayudan en la conceptualización de los problemas. Los *resúmenes visuales y numéricos* también desempeñan un papel importante en el análisis estadístico. Algunos usos comunes de las tablas o resúmenes en las empresas son en el balance (estado activo, pasivo, capital) y el estado de ingresos y gastos que la mayoría de las empresas presenta anualmente.

El propósito primordial de organizar y clasificar los datos estadísticos es permitir visualizar todas las posibles características en los datos que se han recolectados. Se persigue hacer destacar características tales como el rango (los valores mayor y menor), las tendencias aparentes hacia donde tienden a agruparse los datos, los datos que aparecen con mayor frecuencia, el porcentaje de datos en cada grupo de valores, etc.

Entre mayor sea la información que se pueda obtener de la muestra, mejor será la comprensión que se tendrá de la población de la cual proviene y mejores serán las decisiones a tomar.

3. DATOS ESTADÍSTICOS

Se entiende por *datos estadísticos* a las agrupaciones de cualquier número de observaciones relacionadas que miden una característica común de un suceso o evento en una unidad de medida en el tiempo y en el espacio. Esta relación de los datos con el tiempo y el espacio son de mucha importancia estadística, y sin ella es imposible hacer comparaciones entre distribuciones similares que sucedieron o puedan suceder en el futuro.

Así, si se recopilan los datos correspondientes a las exportaciones de un producto comercial de un país a otro, y se dan los datos en tablas y gráficas, pero se olvida mencionar, el año, el mes, o el tiempo de duración de estas exportaciones, entonces esta información carece de valor para un futuro, puesto que no se puede comparar con otro período en que se produce esta actividad comercial.

4. CLASES DE DATOS

Las variables se clasifican en dos tipos: *cuantitativas y cualitativas*. En la figura 1 se da un esquema de la clasificación de las variables:

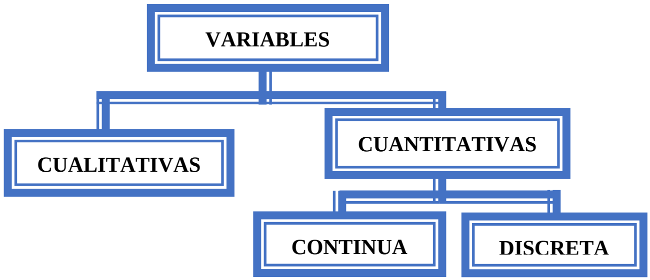

Figura 1. *Clasificación de las Variables.*
Fuente: Elaboración propia.

a) Variable Cuantitativa

Si los posibles resultados de una variable se pueden expresar numéricamente, es una *variable cuantitativa.*

Ejemplo:

El peso de los estudiantes en kilogramos, el número de receptores de televisión de un distrito, los kilómetros recorridos por un automóvil en *x* tiempo, etc. En todos estos ejemplos el valor de la variable cuantitativamente puede expresarse numéricamente. Ahora, si la variable cuantitativa puede asumir cualquier valor en determinado intervalo de valores, entonces se denomina *Variable Cuantitativa Continua.*

Características medibles tales como la estatura, fecha, longitud, espesor, velocidad, viscosidad, ingresos, temperatura, distancia, serán variables de tipo continua. Los datos que se toman acerca de estas características y otros similares se conocen como *datos continuos* (Walpole, Myers y Myers, 1999).

Ejemplo:

La velocidad de un proyectil, tiempo empleado en ejecutar una obra, cantidad de café (lb.) vendidos cada día, gasolina (galones) que se expende por hora, etc.

Una variable cuantitativa, también puede tomar solo valores enteros en tal caso se le denomina de tipo *discreto*. Los datos discretos surgen al contar el número de elemento que poseen ciertas características de las variables.

Ejemplo:

Número de hijos por familia, cantidad de estudiantes en un salón de clases, defectos mecánicos de un automóvil, el número de accidentes de trabajo en una fábrica, las paradas de un autobús, el número de libros de una biblioteca, etc.

Todos estos ejemplos se expresan en valores enteros, y no admiten racionamiento. Por tanto, no tiene sentido expresar que una familia tuviera 3.5 hijos; o que en un salón de clases hay 2.5 estudiantes.

b) Variable Cualitativa

El otro tipo de variable es *cualitativa* y como su mismo nombre lo indica es aquella variable que expresa la calidad o atributos de un evento o su sexo. No se puede expresar en forma numérica, sino por clases o categorías (Walpole, Myers y Myers, 1999).

Ejemplo:

Un ejemplo de variables cualitativas son el sexo, la nacionalidad, las referencias de marcas, profesiones elegidas, clases de empate de un producto, afiliación política, color de los ojos, etc.

A los datos correspondientes a las variables cualitativas, se les debe convertir en valores numéricos para poder aplicar los diversos métodos estadísticos y someterlos al análisis.

5. VENTAJAS DE LA MUESTRA

1. Mediante una muestra se puede reducir el costo sin sacrificar precisión en la información.
2. Es más fácil controlar los errores en las muestras que en la población.
3. Las muestras proporcionan más rápidamente información.
4. Muchas veces el estudio de todos los elementos de la población implica la destrucción de la misma.

6. RECOLECCIÓN DE DATOS

a) Muestreo aleatorio simple

La técnica más común consiste en utilizar el sistema aleatorio o de loterías. La cual consiste en que a cada elemento de la población tiene la misma probabilidad de ser seleccionada para la muestra.

Se le puede asignar una tarjeta a cada uno de los elementos de la población, echarlas a un recipiente y extraer al azar (aleatoriamente) un número determinado de tarjetas que constituyen las tarjetas asignadas a los elementos de la muestra.

b) Muestreo estratificado

Está técnica hace una partición de la población en subconjuntos, luego selecciona de cada subconjunto un número de elementos o sujetos, usando para ello un procedimiento aleatorio. El conjunto de todos los elementos o sujetos seleccionados es esta forma es una muestra por conglomerado.

c) Muestreo deliberado

Esta técnica de muestreo voluntario, pues elige como muestra una parte de la población que nos parece provechoso para la investigación, porque permite poder obtener fácilmente la información. Es un muestreo por conveniencia, comodidad, etc., por lo tanto, no es aleatorio necesariamente.

Ejemplo 1

Está investigando la opinión de su curso respecto a un profesor *X*, en el sentido de sí es bueno o malo. No quiere preguntarles a todos los estudiantes, pero quiere asegurarse que la opinión refleje el pensamiento del curso. ¿Cuál es la población? ¿Cuál es la muestra y cómo la selecciona? ¿Qué método de muestreo utiliza?

Solución:

La población, es el conjunto de estudiantes del curso; y el tamaño de la muestra mientras más grande, las conclusiones se aproximan más a las conclusiones de la población. Se elije, por ejemplo, una muestra que tenga la mitad del número de estudiantes del curso.

En cuanto a la muestra misma, se colocan en una bolsa tarjetas del mismo tamaño en las cuales estan impresos los nombres de todos los estudiantes del curso y se selecciona (aleatoriamente) un número de tarjetas igual a la mitad de los estudiantes del curso. O se utiliza un programa de estadística en el computador, como Excel, que tiene dos funciones que dan números aleatorios, las cuales son de mucha utilidad para seleccionar la muestra.

La muestra es el conjunto de estudiantes correspondientes a las tarjetas seleccionadas o los que en Excel obtuvieron un número aleatorio que cumpla con la condición de selección puesta inicialmente, por ejemplo, seleccionar a los estudiantes que sacaron un numero aleatorio mayor que 0.5.

El método utilizado, en este caso, ha sido el de muestreo aleatorio simple.

Ejemplo 2

¿Cómo se elige la muestra si se utiliza el método de muestreo por estratificación?

Solución:

Una manera conveniente puede ser escribir una lista de todos los estudiantes del curso en orden alfabético y clasificarlos por la primera letra del apellido. Así se ha particionado la población.

De cada grupo ordenado alfabéticamente se saca, al azar, un número determinado de nombres. Por ejemplo, se sacan 3 de las letras A, B, y así sucesivamente. Los elegidos así constituyen una muestra por conglomerado.

Ejemplo 3

¿Cómo se elige la muestra si se utiliza el método de muestreo deliberado?

Es importante resaltar que las muestras deliberadas no tienen cabida en la inferencia estadística.

Solución:

Se conoce la opinión de los amigos del curso, pero no la opinión de los que son simplemente conocidos. Una manera conveniente es elegir como muestra a todos los estudiantes que son simplemente conocidos. Aquí, deliberadamente, se necesita averiguar la opinión de aquel sector y se ha elegido una muestra deliberada.

Una vez elegida la muestra, es necesario asignarle a cada elemento de la muestra el dato correspondiente de acuerdo con la investigación que se está realizando para establecer el conjunto de datos.

Ejemplo 4

En este ejemplo el conjunto de datos es: Bueno, Malo (estos son datos cualitativos). Puede empezarse por hacer un listado de los elementos de la muestra y a cada elemento de ella se le asigna el dato correspondiente. En este ejemplo podríamos escribir tal como se aprecia en la tabla 1

Tabla 1.
Datos obtenidos respecto a la opinión del profesor X

Numero de Estudiantes	Opinión sobre el Profesor
ESTUDIANTE 1	Bueno
ESTUDIANTE 2	Bueno
ESTUDIANTE 3	Malo
ESTUDIANTE 4	Bueno
ESTUDIANTE 5	Malo

Fuente: Elaboración propia.

7. EJERCICIOS RESUELTOS

- *Ejemplo 1*

Determinar cuales de las siguientes variables corresponden a una variable de tipo discreta o si son una variable de tipo continuo.

a) Número de hijos por familias
b) La temperatura
c) Número de viviendas
d) Horas empleadas en ejecutar un trabajo
e) Número de enfermedades infectocontagiosas
f) El peso de una persona

a) **variable discreta**
b) **variable continúa**
c) **variable discreta**
d) **variable continua**
e) **variable discreta**
f) **variable continúa**

- *Ejemplo 2*

Determine las diferencias que hay entre datos cuantitativos y cualitativos y de ejemplos.

La diferencia que existe entre estos datos es que los cuantitativos me permiten expresar los resultados de una variable en forma numérica; mientras que los cualitativos son los que me permiten expresar una cualidad de un evento o suceso.

CUANTITATIVOS (» El tiempo. » El número de hijos. » El número de estudiantes. » El número de accidentes. » El número de artículo de una referencia. » El peso de una persona. » La temperatura. » El número de empleados de una empresa. » La velocidad.)

CUALITATIVOS (» El sexo » La nacionalidad » Los colores » La edad » la profesión » La raza » El estrato social » La religión » El estrato social)

- *Ejemplo 3*

Se realizar en hacer una investigación sobre las estaturas de todos los estudiantes un colegio. ¿Cuál es la población? ¿Cuál es una muestra de la población?

Solución:

La investigación se lleva a cabo sobre todos los estudiantes del colegio A, entonces:

P = Población = {*x: x* es estudiante del colegio A}

Donde P es la población del conjunto de todos los estudiantes del colegio A. Una muestra M, es cualquier subconjunto de la población P. Y x es la muestra del conjunto formado por todos los estudiantes de un curso.

- *Ejemplo 4*

Va a realizar un estudio sobre la contaminación ambiental de todas las ciudades de más de 500.000

habitantes en Colombia. ¿Cuál es la población? De un ejemplo de la muestra.

Solución

La población es el conjunto de todas las ciudades de Colombia que tienen más de 500.000 habitantes, luego:

P = {x: x es ciudad de Colombia con más de 500.000 habitantes} = {Bogotá, Medellín, Cali, Barranquilla}

Una muestra puede ser: M = {Medellín, Cali}.

- *Ejemplo 5*

Usted desea conocer algo sobre el color del cabello de cada uno de los estudiantes del colegio.

Solución

La población es el conjunto de todos los estudiantes del colegio. Una muestra puede ser el conjunto de estudiantes de un curso.

Los datos son los colores de los cabellos de todos los estudiantes del colegio. Si en el colegio solamente hay estudiantes de cabello negro y rubio, entonces el conjunto de datos es: {negro, rubio}.

Una variable aquí sería C, que representa el color del cabello de los estudiantes, esto es negro, o rubio.

Observación:

Es recomendable tomar como mínimo el 30% de la muestra para que el estudio que se esté realizando sea bastante confiable.

8. DISTRIBUCIONES DE FRECUENCIA

La distribución de frecuencia permite clasificar los datos estadísticos en intervalos o clases, para así poder establecer la cantidad de datos presentes en dicho intervalo, esta cantidad puede ser expresada en porcentaje (es decir, la frecuencia) en cada clase. Ordenar los datos de una variable de esta forma permite observar una gran cantidad de datos sin que tenga que considerar cada uno, esto es útil para analizar y manejar grandes cantidades de información. La cantidad de datos o porcentaje en cada clase o intervalo se denomina "*frecuencia de clase*".

Tabla 2.
Distribución Frecuencia de las Edades de los obreros

Edades (Años)	**Frecuencia**	**Porcentaje**
20 - 25	53	0.11
25 - 30	129	0.27
30 - 35	125	0.26
35 - 40	91	0.19
40 - 45	57	0.12
45 - 50	24	0.05
	479	1.00

Fuente: Elaboración propia.

Lo primero que hay que definir para la construcción de una distribución de frecuencia es el número de intervalos o clases se van a utilizar, una vez definido esto se precede a calcular la amplitud va a tener cada uno de ellos. Los pasos anteriores implican que la persona o estadístico conozca a cabalidad las características generales de los datos y el uso adecuado que les vas a dar. Esa persona es el propio investigador y quien está manipulando los datos.

Se necesita un esfuerzo para obtener la mayor claridad en la presentación de los datos sin oscurecer ningunos de los rasgos sobresalientes. Muy pocos intervalos pueden dar como resultado una pérdida de detalles, y tomar muchos puede no condensar satisfactoriamente los datos originales. En términos generales, para grandes cantidades de datos se requieren más intervalos de clases que para pequeñas cantidades de datos.

9. REGLAS GENERALES PARA ORGANIZAR DATOS EN TABLAS DE FRECUENCIA

Siempre deben tenerse en cuenta las siguientes reglas generales para la organización de los datos o información recolectada en las tablas de frecuencia:

Agruparse la información entre cinco a quince intervalos o clases.

Seleccionar este número de intervalos facilita balancear la cantidad y el resumen de los datos y la información que se pierde en el proceso. Cuando se aumenta la amplitud de cada clase, se disipan detalles. Pero entre más intervalos de clases se tomen, más dificulta la extracción de información útil en la tabla.

Se sugiere tomar clases de igual amplitud.

Los intervalos de clases de igual amplitud tienen a dar más forma a las comparaciones. Esto implica la construcción de histogramas más claros, lo que facilita comprender el área y las clases, además basta solo con mirar las alturas de los rectángulos para comparar las clases.

La relación entre el número de clases de una distribución de frecuencia y la amplitud de cada clase es inversa. Ya que se toman menos intervalos de clases, entonces la amplitud aumenta. Por el contrario, si se toma más intervalos de clases, entonces la amplitud disminuye. La siguiente es la relación que existe entre dos (2) conceptos:

$$Amplitud\ de\ los\ \mathrm{int}ervalo\ de\ clase\ = \frac{Dato\ mayor\ \text{-}\ Dato\ menor}{Número\ \ de\ \ clase}$$

Se observa que en la Tabla 2.

Distribución Frecuencia de las Edades de los *obreros* tiene una amplitud de intervalo igual a 5, y se ha utilizado 6 clases.

Una vez que sean definido el número de clases a utilizar, y se comienza la amplitud de cada intervalo, se procede a tabular los datos.

O sea, al contar el número de elementos que corresponden a cada clase asignándoles de esta manera su correspondiente frecuencia.

Ejemplo 1.

Una muestra de 150 estudiantes de la Institución Educativa *X*, realizaron un test para medir su cociente intelectual (CI), se desea condensar esta información en una tabla de distribución de frecuencia. Los resultados obtenidos por los estudiantes fueron:

Tabla 3.
Puntajes del C.I. de 150 estudiantes de Tercer Grado

88	91	104	113	125	101	114	105	101	88	126	118	100	111	125	109
119	91	106	120	129	120	109	104	112	101	113	100	106	105	121	128
93	89	124	96	105	95	91	106	93	88	89	100	115	98	108	88
99	120	101	108	118	118	113	114	109	91	104	109	110	113	119	119
106	106	97	104	105	122	112	124	108	121	96	97	99	101	116	118
102	127	121	116	100	95	89	103	115	113	129	91	85	108	103	116
108	98	108	114	102	96	99	108	114	121	107	122	100	116	111	113
109	104	113	118	110	129	124	105	93	115	120	97	102	94	103	122
114	106	105	115	98	112	103	92	125	107	115	118	128	92	85	126
108	114	125	121	122	117										

Fuente: Elaboración propia.

El primer paso que haremos es ordenar los datos de menor a mayor.

Tabla 4.
Puntajes Ordenados del C.I. de 150 estudiantes de Tercer Grado.

85	91	95	98	101	103	105	108	109	113	114	116	119	121	125
85	91	95	99	101	104	105	108	109	113	114	116	119	122	125
88	91	96	99	101	104	106	108	109	113	114	117	120	122	126
88	91	96	99	101	104	106	108	110	113	115	118	120	122	126
88	92	96	100	102	104	106	108	110	113	115	118	120	122	127
88	92	97	100	102	104	106	108	111	113	115	118	120	124	128
89	93	97	100	102	105	106	108	111	113	115	118	121	124	128
89	93	97	100	103	105	106	108	112	114	115	118	121	124	129
89	93	98	100	103	105	107	109	112	114	116	118	121	125	129
91	94	98	101	103	105	107	109	112	114	116	119	121	125	129

Fuente: Elaboración propia.

Tamaño de la muestra: n = 150

El puntaje mínimo es: 85

El puntaje máximo es: 129

Para calcular el rango o diferencia entre el puntaje mayor y el menor se procede de la siguiente manera:

primero se identifica el puntaje más bajo, que es 85 y el más alto que es 129. Y se hace la diferencia, esto es:

El Rango = Máximo – Mínimo = 129 - 85 = 44

Para agrupar estos datos en la tabla de distribución de frecuencia, es necesario determinar el *NÚMEROS DE INTERVALOS DE CLASE* (k). Para obtener un valor aproximado, podemos emplear la regla de ¨*STURGES*¨.

$$k = 1 + 3.3\log(n)$$

Donde n es el número de elementos de la muestra. Apliquemos esto a la muestra de 150 estudiantes. Es decir $n = 150$

$$k = 1 + 3.3\log(n) = 1 + 3.3\log(150) = 8{,}18 \cong 9$$

Con 9 intervalos o clases, entonces, la amplitud del intervalo se calcula, así:

$$I = \frac{rango}{número\ \ de\ \ clases} = \frac{44}{9} = 4{,}88 \cong 5$$

En la práctica, el valor decimal se redondea al valor entre, lo que resulta que usaremos un intervalo *igual a 5*.

Con base en esta información y requerimiento construyamos una tabla de frecuencia para los datos de la tabla anterior.

Tabla 5.
Distribución de Frecuencia y de Frecuencia Relativa de los Puntajes del C.I. de los 150 estudiantes

Núm. de Intervalos	Intervalos		Marca de clase (xi)	Frecuencia Núm. de datos (fi)	Frecuencia relativa %. (hi = fi/n * 100%)	Frecuencia acumulada de Núm. de datos (Fi = fi + fi + 1)	Frecuencia relativa acumulada (Hi = Fi/n * 100%)
	Límite Inferior (Li)	Limite Superior (Ls)					
1	85	90	87,5	9	6%	9	6%
2	90,1	95	92,5	13	9%	22	15%
3	95,1	100	97,5	17	11%	39	26%
4	100,1	105	102,5	23	15%	62	41%
5	105,1	110	107,5	23	15%	85	57%
6	110,1	115	112,5	23	15%	108	72%
7	115,1	120	117,5	18	12%	126	84%
8	120,1	125	122,5	16	11%	142	95%
9	125,1	130	127,5	8	5%	150	100%
	Sumatoria			150	100%		

Fuente: Elaboración propia.

Analizando la tabla 5 podemos determinar rápidamente el número de observaciones que se incluyen en las diferentes clases o intervalos de clases. Así, por ejemplo, vemos que hay 17 estudiantes que obtuvieron un cociente intelectual entre 95 y 100 incluso. También se observa que 23 estudiantes obtuvieron un puntaje entre 105,1 y 110 inclusive.

Las frecuencias relativas o porcentajes lo constituyen la relación de elemento de cada clase con respecto al total de estudiantes.

Así, se observa que un 15% de los estudiantes tienen un porcentaje entre 105,1 y 110 incluso. Esto es:

$$hi = \frac{fi}{n} = \frac{23}{150} = 0{,}15 = 15\%$$

Comúnmente las frecuencias se representan con el símbolo *"f_i"*.

Al igual que las frecuencias relativas o porcentuales se simbolizan con *"h_i"*, así que:

$$\text{Frecuencia porcentual } \mathrm{h_i} = \frac{f_i}{n} = \text{"\%"}$$

Analizando el puntaje del cociente intelectual de los 150 estudiantes, por medio de la distribución de frecuencia en la tabla 3 se puede responder de una manera más concisa, las siguientes preguntas:

a. ¿Cuántos estudiantes obtuvieron un puntaje superior a 110 puntos?
b. ¿Qué porcentaje de estudiantes obtuvieron un puntaje menor de 95 puntos?
c. ¿Cuántos estudiantes obtuvieron un puntaje comprometido entre 110 y 125 puntos inclusive?
d. ¿Qué porcentaje de estudiantes obtuvieron notas por encima de 115 puntos? Y, ¿cuántos estudiantes representan este porcentaje?
e. ¿Cuántos estudiantes, y que porcentajes representan aquellos que obtuvieron un puntaje por encima de 100 puntos, pero por debajo de 120 puntos?

Solución:

Observando las frecuencias obtenidas en la tabla 3, podemos obtener las respuestas.

Puntaje superior a 110 puntos:
Puntaje entre 110,1 y 115 = 23 estudiantes
Puntaje entre 115,1 y 120 = 18 estudiantes
Puntaje entre 120,1 y 125 = 16 estudiantes
Puntaje entre 125,1 y 130 = 8 estudiantes
Total = 65 estudiantes

Hay 65 estudiantes que obtuvieron puntajes superiores a 110 puntos. Otra manera de resolver esta pregunta es analizando las frecuencias acumuladas *"F_i"*.

a. No. Estudiantes con puntajes a 110 puntos:

Total, de estudiantes - el No. estudiantes con el porcentaje por debajo de 110 puntos:

150 - 81 = 69 estudiantes

b. % de estudiantes con puntajes menores de 95:

9 + 13 = 22 estudiantes $\frac{22}{150} = 0{,}15 = 15\%$

** También, se puede leer directamente en la tabla, en la columna de la frecuencia acumulada porcentual = 15 %.

c. No de estudiantes: 23+18+16 = 57 estudiantes.

d. No de estudiantes: 18+16+8= 42 estudiantes. Porcentaje de estudiante.

$$\% \text{ de estudiantes}: \frac{42}{150} = 0{,}28 = 28\%$$

e. No de estudiantes: 23+23+23+18 = 87 estudiantes. Porcentaje de estudiantes: 87/150 = 58 %.

En la tabla 3 se han incluido los límites reales de cada clase, que son los valores a los cuales se han aproximado o redondeados los datos con el objeto de incluir todos los valores continuos de las variables.

Ejemplo 2

Sea *Xi* la estatura (en pulgadas) de 50 estudiantes de un colegio.

Tabla 6.
Datos de las estaturas de los estudiantes

65	64	64	63	64
67	65	65	65	64
63	64	64	63	63
53	72	71	70	69
65	68	68	67	67
58	66	66	66	66
63	55	56	57	58
60	57	59	59	60
69	60	61	61	61
61	62	62	62	62

Fuente: Elaboración propia.

Tabulación de los datos de las estaturas de los 50 estudiantes:

Tabla 7.
Tabulación de Datos

Estaturas (pulgadas)	Puntos Medios	Conteo	Número de estudiantes
50.5-53.5	52	I	1
53.5-56.5	55	II	2
56.5-59.5	58	IIIII	6
59.5-62.5	61	IIII IIIII	11
62.5-65.5	64	IIII IIII IIIII	16
65.5-68.5	67	IIII IIII	9
68.5-71.5	70	IIII	4
71.5-74.5	73	I	1
			50

Fuente: Elaboración propia.

Con el objeto de ordenar forma más sistemática los datos de la variable X relacionados con las estaturas de los 50 estudiantes, es preciso elaborar una tabla de distribución de frecuencia, definiendo *frecuencia asociada* de una clase determinada como el número de veces que se presenta el valor del punto medio de cada clase.

Este punto medio se suele designar como la *marca de clase* y se simboliza por x_i y las *frecuencias asociadas* a cada clase se designa por f_i. De igual manera se designa por *hi* a las *frecuencias relativas,* es decir, la proporción de veces que se toma el valor *xi*.

Se consideran además las *frecuencias acumuladas* (*Fi*) que representan la acumulación sucesiva de cada clase hasta agrupar todas las observaciones ($f_1 + f_2 + f_3 + \dots f_n$). Análogamente, *Hi* representa las *frecuencias relativas* acumuladas de cada clase. Esto es, *Hi* es la proporción de veces que la variable *Xi* toma un valor que no exceda a *xi*.

Tabla 8.
Distribución de Frecuencia-Estatura 50 Estudiantes

Intervalos Li-Ls	x_i	f_i	Fi	h_i	H_i
1.4-53.5	52	1	1	0.02	0.02
1.5-56.5	55	2	3	0.04	0.00
1.6-59.5	58	6	9	0.02	0.18
1.7-62.5	61	11	20	0.22	0.40
1.8-65.5	64	16	36	0.32	0.72
1.9-68.5	67	9	45	0.18	0.90
1.10-71.5	70	4	49	0.08	0.98
71.5-74.5	73	1	50	0.02	1.00
		50		**1.00**	

Fuente: Elaboración propia.

10. PROPIEDADES Y RELACIONES DE LAS DISTRIBUCIONES DE FRECUENCIAS

Las propiedades más importantes de las distribuciones de frecuencias las podemos expresar así:

La frecuencia sencilla (f_i), al igual que las frecuencias acumuladas (F_i), son números enteros no negativos.

$$f_i \geq 0 \qquad F_i \geq 0$$

Las frecuencias relativas (h_i), y las frecuencias relativas acumuladas (H_I) son números no negativos ni mayores de 1.

$$0 \leq h_i \leq 1 \qquad \sum_{i=1}^{n} h_i = 1$$

Las frecuencias más importantes se pueden expresar como.

1. $\sum_{i=1}^{m} f_i = n$
2. $\sum_{i=1}^{m} f_i = Fm$
3. $\sum_{i=1}^{m} h_i = Hm$
4. $Fm = n$
5. $\sum_{i=1}^{m} h_i = 1$
6. $H_m = 1$
7. $f_i = F_1 \leq F_2 \leq F_3 \ldots \leq F_m = n$
8. $h_i = H_1 \leq H_2 \leq H_3 \ldots \leq H_m = 1$

11. REPRESENTACIÓN GRÁFICA DE LAS DISTRIBUCIONES DE FRECUENCIA

Es de mucha utilidad e importancia destacar la información procesada en una la tabla de distribución de frecuencia. Es por ello que la tabla o cuadro de resúmenes de los datos estén siempre acompañados con sus respectivas gráficas. Existen una gran variedad de formas para representar los datos en forma gráfica. La variedad de forma y disposición de las gráficas depende del tipo variables: *discretas o continuas*.

Generalmente se representan mediante gráficos, tales como diagramas de barra o histogramas (histogramas para variables continuas y barras para discretas), polígonos de frecuencias relativas y acumuladas, gráficos circulares o de torta u otras formas. Al construir la gráfica, los intervalos de clase se ubican en el eje de las abscisas o eje X, y la frecuencia se representan por el eje de las ordenada o eje *Y*.

El diagrama de Barras o Histograma

Un Histograma es una serie de rectángulos, cada uno proporcional en amplitud al rango de valores dentro de una clase y proporcional en altura al número de elementos que posee cada clase.

Si las clases que se usan en la distribución de frecuencia son de igual amplitud, entonces las barras verticales en el histograma serán también de igual amplitud. La altura de cada rectángulo en cada clase corresponde al número de elementos en cada clase. Como resultado el área contenida en cada rectángulo (base x=altura), es el área total de todos los rectángulos en porcentaje, como la frecuencia relativa de esa clase es el total de las observaciones obtenidas.

En resumen, *los histogramas* presentan las siguientes características:

1. Los posibles valores de las variables que estan graficando se disponen en el eje horizontal. La frecuencia absoluta "f_i" con que ocurren los valores de las variables se representan en el eje vertical.
2. Los intervalos de clase de la distribución de frecuencia se representan por una barra del histograma.
3. La barra o rectángulo tiene la misma amplitud de los intervalos de clase correspondiente.
4. La altura de las barra corresponden a la frecuencia con que ocurren los valores en el correspondiente intervalo de clase.

El Polígono de Frecuencia

Los datos de distribución de frecuencia se pueden describir por medio de otra representación gráfica la cual recibe el nombre de "*polígono de frecuencia*".

Para construir fácilmente un polígono de frecuencia se dibuja primero un histograma y luego se conecta por medio de líneas rectas, los puntos medios de la parte superior de cada una de las barras o rectángulo.

El boceto del histograma generalmente se suprime en los polígonos de frecuencia.

Observase que en el eje vertical o eje de la Y (o altura) se pueden utilizar instintivamente ya sean frecuencias absolutas (f_i) o la frecuencia relativa (h_i).

Algunas características importantes del Polígono de Frecuencia son las siguientes:

1. El polígono se construye uniendo con segmentos consecutivamente los puntos medios del lado superior de cada rectángulo del histograma.
2. El área bajo la curva el polígono es igual al área del histograma de frecuencia.

12. TÉCNICAS DE REPRESENTACIÓN GRÁFICA

Recordemos que la información estadística se presenta en forma numérica. Si la tabla es muy extensa puede ser difícil alcanzar una comprensión clara de la información presentada al igual que sus características. Para lograr una mayor comprensión de los números arreglados en forma tabular utilizamos los gráficos que nos destaquen algunos hechos más claramente.

Un gráfico para ser de utilidad real, debe ser simple y poner mayor énfasis en los rasgos más significativos de los datos. Por lo tanto, el siguiente paso después de clasificar los datos es la representación de los datos en forma gráfica, de tal forma que el lector pueda percibir fácilmente los hechos esenciales de una distribución de frecuencia y compararlos con otra, si lo desea. Son esencialmente una ayuda visual para pensar e interpretar problemas estadísticos.

Representación de escalas variables nominales

La representación gráfica de las escalas nominales y ordinales se realiza a través del diagrama de barras; para cada categoría se traza una barra vertical en la que la altura de la barra representa el número de miembros de esa clase. Si se fija arbitrariamente el ancho de la barra como una unidad, el área de cada barra puede usarse para representar la frecuencia de cada categoría. De esta forma el área total de todas las barras es igual *N*. Para la representación gráfica se debe tener en cuenta:

1. No existe ningún orden para colocar las escalas de las variables nominales, así mismo las categorías pueden ser presentadas a lo largo de las abscisas en cualquier orden. Colocadas en orden alfabético para obviar implicaciones de tipo personal.
2. Es preferible que exista separación y ninguna proximidad de las barras, para así evitar la implicación de continuidad entre las categorías.

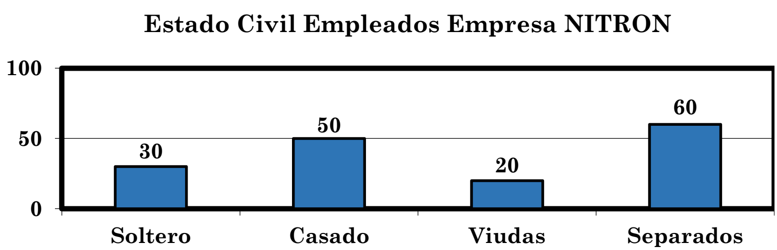

Figura 2. Histograma de frecuencia
Fuente: Elaboración propia.

El uso de la regla de tres cuartos (convención usada por los estadísticos) es perfectamente aplicable a la gráficas de escalas nominales. Esta regla señala:

> *La representación gráfica del histograma de frecuencias, implica que el eje vertical debe tener como máximo al rectángulo de mayor altura que está asociado con la frecuencia más alta, y como mínimo al rectángulo de menor altura que está asociado con la frecuencia más baja.*

Para realizar la gráfica se debe tener en cuenta, por lo tanto, el espacio que se tiene disponible para la representación y de acuerdo a este trazar las abscisas respetando la convención señalada.

Representación de escalas variables ordinales

Los valores de las escalas ordinales implican un ordenamiento que es expresable en términos algebraicos de desigualdades (mayor que, menor que), por lo tanto, las variables ordinales, pueden ser representadas con el diagrama de barras, teniendo en cuenta que sus categorías deben colocarse a lo largo de la abscisa, en el orden natural de ocurrencia.

La figura 3 muestra el uso del diagrama de barras en una escala de variables ordinales que presentan gráficamente los datos de los ganadores de una carrera de acuerdo a las poblaciones de salida.

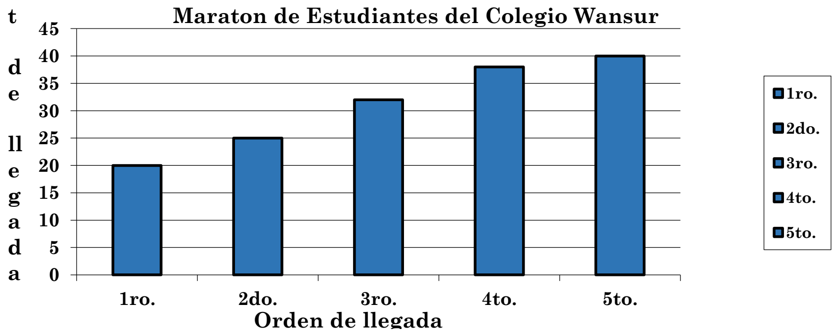

Figura 3. Diagrama de ganadores en carrera
Fuente: Elaboración propia.

Representación de escalas variables de intervalos y de cocientes

- *Histogramas*

Para la representación de variables que utilizan escalas de intervalo y de razón debe tenerse en cuenta sus propiedades. En una escala de intervalos o de razón las diferencias iguales entre dos puntos en cualquier parte de la escala, son iguales entre sí. De ahí que las representaciones gráficas de las escalas de distribución de frecuencias, intervalos o de razón de las barras verticales se toquen unas con otras. Cuando esto sucede el diagrama cambia su nombre por *histograma*. La figura 4 ilustra el empleo del histograma con la escala de razón variable en distribuciones discretas.

Se señaló anteriormente que la frecuencia puede representarse por el área de la barra o también por la altura. Sin embargo, en ocasiones en las cuales la altura de la barra o la ordenada, pueden causar una interpretación equivoca de la información concerniente a la frecuencia, por ejemplo cuando se utilizan intervalos de clases desiguales, se aconsejable que la frecuencia se considere en términos de área cuando se trate de variables en las cuales se puede suponer continuidad.

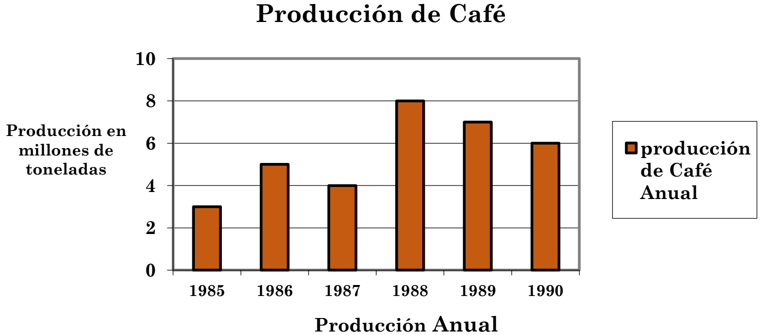

Figura 4. Histograma de producción de café
Fuente: Elaboración propia.

- *Polígono de Frecuencia*

El histograma se puede transformar en una forma empleada comúnmente para la representación gráfica llamada *polígono de frecuencia*, el cual es utilizado fundamentalmente para representar fenómenos que ocurren en períodos sucesivos. El polígono de frecuencia se logra mediante la unión de los puntos medios de las barras a base de segmentos de rectas. Sin embargo, no es necesario elaborar un histograma previo a la construcción del polígono, es suficiente con marcar el punto en el sitio correspondiente a las cimas de las barras y luego unir esos puntos (Figura 5).

En la práctica, los autores prefieren aplicar histogramas a las distribuciones discretas, y el polígono de frecuencia a distribuciones en las cuales la continuidad es explícita o puede ser supuesta. Cuando dos o más distribuciones se comparan entres sí, el polígono de frecuencia proporciona una visión más clara. La figura 5 muestra un polígono de frecuencia en base a la producción de café en millones de toneladas durante el período 1985–1990

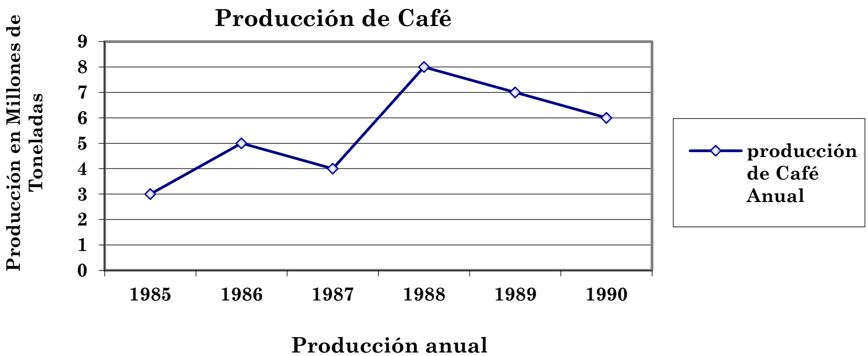

Figura 5. Polígono de frecuencia de la producción de café
Fuente: Elaboración propia.

Cada punto del gráfico representa un par ordenado de valores para la tabla correspondiente; así para 1988 de obtuvo el mayor crecimiento en la producción de café en millones de toneladas.

- *Ojivas*

Una distribución de frecuencia acumulada permite observar cuantas observaciones se encuentran por encima o por debajo de ciertos valores, en lugar de un recuento de los elementos dentro de los intervalos de clase. Si por *ejemplo*, se quiere saber cuántos de los 50 estudiantes referentes a sus estaturas poseen menos de 64.5 pulgadas, nos referimos a una tabla de distribución de frecuencias acumuladas "menor que", tal como aparece en la tabla 9.

Tabla 9.
Distribución de Frecuencia Acumulada en las Estaturas (en pulgadas) de 50 Estudiantes

Estatura (pulgada) Li	Puntos medios Xi	Números de estudiantes F_i	Números de estudiantes (menor que) F_I	Números de estudiantes (mayor que)
50.5 - 53.5	52	1	1	50
53.5 - 56.5	55	2	3	49
56.5 - 59.5	58	6	9	47
59.5 - 62.5	61	11	20	41
62.5 - 65.5	64	16	36	30
65.5 - 68.5	67	9	45	14
68.5 - 71.5	70	4	49	5
71.5 - 74.5	73	1	50	1
		50		

Fuente: Elaboración propia.

Para responder a las inquietudes planteadas, observemos en la ojiva "menor que ", que existen 36 estudiantes con estaturas por debajo de 65.5 pulgadas.

Para responder a la segunda pregunta, se observa en la ojiva “mayor que”, y encontramos que hay 14 estudiantes con estaturas por encima de 65.5 pulgadas.

También al igual que en los histogramas, la escala vertical, que indica la frecuencia se puede utilizar la frecuencia relativa acumulada (H_i)

Representación en Diagramas de Frecuencia (Pastel o Círculo)

Los diagramas de frecuencia llamados también *diagramas de pastel*, se emplean generalmente para representar distribuciones de razones. El círculo total es dividido en partes con forma de cuña por medio del trazado de radios. En la figura 6 se muestran los datos presentados previamente, dispuestos en un diagrama circular. El círculo representa la suma del conjunto de la distribución de razones (100%). Cada porción representa una razón en la serie.

Para construir un gráfico circular debe tenerse en cuenta los siguientes factores:

a. El área total del circulo corresponde al 100%, por tanto, al 1% le corresponde el 3,6º.
b. Toda cantidad parcial se debe expresar en porcentaje.
c. Asignar a cada porcentaje parcial un sector circular de acuerdo con el ángulo correspondiente a dicho porcentaje.

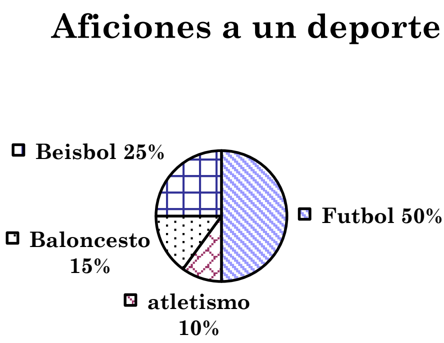

Figura 6. Representación de aficionados a un deporte en forma circular
Fuente: Elaboración propia.

Ejemplo:

Se aplicó una encuesta en el Liceo X a una muestra de 800 estudiantes para identificar cuáles eran sus deportes preferidos; los resultados fueron:

Aficionados al béisbol	200
Aficionados al fútbol	400
Aficionados al baloncesto	120
Aficionados al atletismo	80

Representando la información por un gráfico circular, se obtuvieron las tendencias de los estudiantes del Liceo X a su deporte expresado.

- Determinemos el porcentaje correspondiente a cada deporte:

Aficionados al béisbol	200	porcentaje	25%
Aficionados al fútbol	400	porcentaje	50%
Aficionados al baloncesto	120	porcentaje	15%
Aficionados al atletismo	80	porcentaje	10%

- Teniendo en cuenta que el 1% corresponde a 3,6º de la circunferencia, establecemos una regla de tres simple y hallamos el correspondiente a 25%

Sí 1% equivale a 3,6º
25% equivales a X

$$X = \frac{3{,}6^{o} * 25\%}{1\%} = 90^{o}$$

y así sucesivamente hasta completar 360º el cual conforma una circunferencia. De acuerdo al interés del científico pueden surgir otras formas de gráficas, pero en general las acá explicadas constituyen las formas más generalizadas de representación gráfica estadística.

Otros tipos de Gráficos

Hay algunos datos, que por su simplicidad son fáciles de expresar en forma de gráficos. Tal es el hecho de los Datos nominales y los jerarquizados por categorías. Esta simplicidad radica en el hecho de que las clases que los constituyen se ponen de manifiesto con mucha más facilidad, de tal manera que los cálculos son mínimos.

Por ejemplo, se consideran los datos nominales de la tabla 10, los cuales representan las ventas de un producto comercial ordenadas en una tabla de frecuencia.

Tabla 10.
Ventas de Gaseosas en un establecimiento

Tipos de gaseosa (sabor)	Frecuencias	
	Ventas reales	Ventas relativas
	f_i	% (h_i)
Kola	600	60%
Naranja	200	20%
Limón	100	10%
Uva	50	5%
Fresa	40	4%
Otros	10	1%
	1000	100%

Fuente: Elaboración propia.

Se observa que las categorías en este ejemplo son los diversos sabores de gaseosa. La representación gráfica se hace mediante una gráfica de barra en donde las barras no se tocan, como en los histogramas, esta gráfica de barras se puede construir mostrando las barras ya sean en forma horizontal o vertical, como puede hacerse cualquier grafica de una distribución de frecuencia.

Otra gráfica muy utilizada para representar datos estadísticos en tablas de distribución de frecuencias, son las gráficas circulares. Se trata de representar la frecuencia absoluta o relativas a través de un circulo que representa 100% de la frecuencia, y subdividiendo el circulo en parte proporcional a su frecuencia.

13. EJERCICIOS RESUELTOS

1. ¿Qué ventajas poseen los métodos gráficos en el resumen de datos?

   R/TA:

   La ventaja es que permiten reunir toda la información en forma resumida en un solo lugar para poder analizarla, comprender y sacar las conclusiones necesarias de dicha población.

   Las gráficas no solamente sirven como un instrumento de comunicación sino que también ayudan como conceptualización de los problemas.

2. ¿Por qué son útiles las distribuciones de frecuencias?

   R/TA:

   Porque permiten ordenar y clasificar los datos estadísticos en clases o intervalos de tal manera que se pueden visualizar y establecer las características posibles de los datos que se han recolectado, para así poder obtener la cantidad de datos existentes dentro de esa clase y además permiten medir su porcentaje dentro de la población

AUTOEVALUACIÓN

1. Dada las siguientes variables en los ítems (a, b, c, d y e), determine el tipo de variable que es (cualitativa o cuantitativa).

   a. Cantidad de hijos por familias en un barrio.
   b. Costo de las acciones vendidas en una bolsa de valores por semana.
   c. La velocidad de un carro que se desplaza en kilómetros por horas.
   d. Medida en pulgadas de agua lluvia que cae en una región determinada en un tiempo de dos horas.
   e. Número de horas para la realización de un trabajo.

2. Explique y de ejemplo de variables cualitativas y cuantitativas.
3. ¿Para qué se realizan gráficos en un informe estadístico? De ejemplo.
4. ¿Cuál es la importancia de usar las distribuciones de frecuencia?
5. ¿Cuáles son los pasos para elaborar una tabla de distribución de frecuencia?
6. Una empresa procesadora de alimento, ha recopilado los siguientes datos en libras de los mangos vendidos a los clientes en un período de 40 días

700	1100	800	700	1000	1000	600	500
10000	1200	900	900	1100	700	500	600
800	600	1000	1000	1500	700	800	1300
500	500	500	1000	800	1200	900	800
800	600	800	1400	500	500	800	1400

Construya una tabla de distribución de frecuencia, con sus intervalos igualmente espaciados y de una amplitud mínima de 200 libras.

7. En una oficina de prensa, el tiempo requerido para armar la totalidad de la primera página a máquina fue registrado durante 50 años con datos que se aproximan a la décima de un minuto más cercana, y se dan enseguida.

20.8	22.8	21.9	22.0	20.7	25.0	22.2	22.8	20.1
25.3	20.7	22.5	21.2	23.8	20.9	22.9	23.5	19.1
23.7	20.3	23.6	19.0	25.1	19.5	24.1	24.2	21.8
21.3	21.5	23.1	19.9	24.1	19.8	23.9	22.8	23.9
19.7	24.2	23.8	20.7	24.3	21.1	20.9	21.0	22.7

a) Organice los datos de un arreglo de menor a mayor.
b) Construya una distribución de frecuencia y una distribución de frecuencia acumulada “menor que” a partir de los datos utilizando intervalos de 0.8 minutos.
c) Construya un polígono de frecuencia a partir de los datos.
d) Construya una ojiva de frecuencia “menor que” a partir de los datos.

8. El concejo nacional de seguridad hizó un muestreo aleatorio de la profundidad de huella de 60 llantas delanteras derechas pertenecientes a vehículos de pasajeros en una autopista. Con base en los datos, se construyó la siguiente distribución de frecuencia. ¿Cuál es la profundidad de la huella de la 30ª llanta en esta distribución de frecuencia?

Profundidad de la huella (pulgada)	Frecuencias
16/32 llantas nuevas	5
13/32 - 15/32	10
10/ - 12/32	20
7/32 - 9/32	12
4/32 - 6/32	7
1/32 - 3/32	4
0/32 - (lisa)	2

9. Antes de construir una represa los ingenieros militares realizaron una serie de pruebas para medir el flujo de agua que pasaba por la localización propuesta por la represa. Los resultados de la prueba son:

Flujo del rio (miles galones por minutos)	Frecuencias
1.001-1.050	9
1.051-1.100	20
1.101-1.150	31
1.151-1.200	45
1.201-1.250	53
1.251-1.300	39
1.301-1.350	27
1.351-1.400	13
total	237

a) Use los datos de la tabla para construir una distribución de frecuencia acumulada “mayor que” y una ojiva.
b) Use los datos de la tabla para construir una distribución de frecuencia acumulada “menor que” y una ojiva.

10. Dada la siguiente distribución de frecuencias de los pesos de 96 estudiantes de sexo femenino, dibujar una gráfica que los represente:

Intervalo de Clase	Frecuencia (f)
160 – 164	1
155 – 159	3
150 – 154	10
145 – 149	6
140 – 144	14
135 – 139	22
130 – 134	17
125 – 129	11
120 – 124	8
115 – 119	3
110 – 114	1

11. Tomar un par de dados, lanzarlos 50 veces y anotar la suma de los números que aparecen en la cara superior de los dados en cada lanzamiento. Construir una gráfica de barra que muestre la ocurrencia de la cada suma.
12. Los siguientes son los votos obtenidos por cuatro candidatos a la presidencia de una compañía.

Candidato	Voto
A	12
B	16
C	8
D	25

Dibujar una gráfica que represente los resultados de la elección. Reemplazarlo por un diagrama circular.

13. Describir los tipos de distribuciones que obtendrías si tuvieras que representar gráficamente las siguientes situaciones de su Institución.

    a) El ingreso salarial de sus empleados
    b) La altura de los alumnos hombres
    c) La altura de las alumnas mujeres
    d) La altura de los alumnos hombres y mujeres combinados en una sola gráfica.

14. Busca en un periódico o revista algunas gráficas que represen o muestren datos de distribuciones estadísticas. Trate de interpretarlos con un grupo de compañeros y determine cuál tipo de variables o escala están representado.
15. ¿Cuáles otros modelos de representación gráfica de distribuciones estadísticas conocen y cuando se usan?
16. ¿Qué diferencia existe entre un diagrama de barras, un histograma y un polígono de frecuencias?
17. ¿Cuándo se utilizan las gráficas circulares?

CAPITULO III

MEDIDAS DE TENDENCIA CENTRAL
Y MEDIDAS DE VARIABILIDAD

INTRODUCCIÓN

El análisis de datos estadísticos se puede llevar a cabo de diversas maneras dependiendo de la disponibilidad de una gran cantidad de datos, o pocos datos.

Si el número de datos que se van a analizar son pocos, entonces los métodos utilizados serán los que se dan en esta unidad; pero si los datos son bastante numerosos, entonces será necesario el uso de computadoras, para su agrupamiento, antes de proceder al análisis de ellos.

En la mayoría de los casos el conjunto de datos que se está analizando se puede reducir a unas cuantas medidas que resumen el conjunto total haciendo más fácil la comprensión de los datos originales no procesados.

Al analizar este conjunto de datos se procura siempre poner en manifiesto dos importantes características de los datos:

1. Las medidas de tendencias centrales.

2. La dispersión de los datos respecto a estos valores centrales.

La primera característica la constituyen las medidas de tendencia central, tales como la medida, la mediana, la moda, etc. La segunda característica la conforman los valores que expresa una dispersión de los datos tales como la varianza, la desviación estándar y el coeficiente de variación.

1. MEDIDAS DE TENDENCIA CENTRAL

La Medida Aritmética o Promedio

Forma de Cálculo

Sea $X_i = \{X_1, X_2, X_3, \dots X_n\}$, Si sumamos cada uno de los valores de la variable X, y dividimos el número de valores sumados, obtendremos un indicador muy importante en el análisis estadístico de los datos, llamado *medida aritmética* o *promedio*. En símbolos, se expresa tal como sigue:

$$\overline{X} = \frac{X_1 + X_2 + X_3 + X_4 + \cdots X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Ejemplo:

Las edades de 10 personas son: 22, 37, 60, 46, 18, 35, 41, 27, 24 y 40. ¿Cuál es la edad promedio de este grupo de personas?

Edades (años): 22, 37, 60, 46, 18, 35, 41, 27, 24, 40

Sumando:

22+37+60+46+18+35+41+27+24+40 = 350

Dividendo: $\frac{\sum_{i=1}^{n} X_i}{n} = \frac{350}{10} = 35$

Solución: la edad promedio es de 35 años.

Hay situaciones donde los datos de la variable se dan en tablas o cuadros estadísticos por niveles de clase. Entonces, en este caso, es necesario utilizar un modelo de cálculo, un poco diferente.

Cada clase o nivel se repite un número *k* de veces, entonces, su suma se hace con ponderaciones de valores:

Tabla 11.
Distribución de Frecuencias

Niveles o clase	Marca de clase	Frecuencias	$X_i f_i$	Frecuencias porcentuales	$X_i (h_i)$
Edades	X_1	No. de personas		% (h_1)	
5 -9	7	10	70	7.69	0.5383
9 - 13	11	17	187	13.08	1.4388
13 - 17	15	25	375	19.23	2.8845
17 - 21	19	35	665	26.92	5.1148
21 - 25	23	22	506	16.92	3.8916
25 - 29	27	14	378	10.77	2.9079
29 - 33	31	7	217	5.39	1.6709
		130	2.398	100.00	18.446

Fuente: Elaboración propia.

$$\text{Promedio X} = \frac{\sum_{i=1}^{n}(X_i f_i)}{n} = \frac{2398}{130} = 18{,}446$$

La edad promedio de estas 130 personas es de 18,45 años aproximadamente.

Esto se denomina *promedio ponderado*.

Otra manera de calcular la medida aritmética o promedio es cuando los datos están tabulados en un cuadro de frecuencia expresado en forma de porcentaje. Se deberá utilizar la siguiente fórmula:

$$\text{Promedio X} = \sum_{i=1}^{n}(X_i\ h_i)$$

Por medio del cual, obtenemos también un promedio de 18,45 años

Propiedades de la Medida Aritmética

Primera: Si se suman las desviaciones de los datos de una variable con respecto a la medida aritmética y esta es igual a cero.

Se entiende por desviación la diferencia de cada valor con relación al promedio de la variable:

Ejemplo:

Si los ingresos 8 familias se discriminan así:

Miles de pesos: 15, 18.5, 27.3, 25.5, 18.5, 33.2, 29.5, 40.5

Se calcula el promedio:

$$\overline{X} = \frac{15 + 18.5 + 27.3 + 25.5 + 18.5 + 33.2 + 29.5 + 40.5}{8} = 26$$

Calculemos ahora las desviaciones de los datos:

$$(15 - 26) + (18.5 \;\; 26) + (27.3 \;\; 26) + (25.5 \;\; 26) + (18.5 \;\; 26) + (33.2 \;\; 26) + (29.5 \;\; 26) + (40.5 \;\; 26)$$
$$= (-11) + (-7.5) + (1.3) + (-0.5) + (-7.5) + (7.2) + (3.5) + (14.5)$$
$$= (-26.5) + (26.5) = 0$$

Segunda: si todos los valores $X_1 + X_2 + X_3 + X_4 + \cdots X_n$ son iguales a *K* (*K* constantes), entonces el promedio será igual a la misma constante *K*.

Ejemplo:

Si una persona tiene ingresos mensuales de $18.500, entonces su ingreso durante el año será de $18.500.

$$\overline{X} = \frac{18.500 + 18.500 + 18.500 + \ldots..18.500}{12} = \frac{222.000}{12} = 18.500$$

Tercera: Si a cada uno de los datos de una variable se le suma una misma cantidad constante, entonces el promedio será igual al promedio de la variable más constante**.**

$$\overline{X}_n = \frac{(X_1+K)+(X_2+K)+(X_3+K)+(X_4+K)+\cdots(X_n+K)}{n}$$

$$= \frac{(X_1+X_2+X_3+X_4+\cdots X_n)+nK}{n}$$

$$= \frac{\sum_{i=1}^{n} X_i}{n} + \frac{nK}{n}$$

$$= \overline{X} + K$$

Ejemplo:

Si al ejemplo anterior, se le suma a cada salario \$1.500 entonces el nuevo promedio se calcula así:

Salarios (miles de \$): X_i = 15, 18.5, 27.3, 25.5, 18.5, 33.2, 29.5, 40.5

$$(\mathrm{Xi}+1.5) = 16.5+20+28.8+27+20+34.7+31.42$$

$$\overline{X} = \frac{220}{8} = 27.5$$

Aplicando la propiedad:

$$(\mathrm{X}+\mathrm{K}) = 26+1.5 = 27.5$$

En la práctica siempre se debe aplicar la propiedad.

Cuarta: Si cada uno de los datos de una variable se multiplica por una constante entonces el promedio será igual al promedio de la variable por la constante, es decir el nuevo promedio se amplifica por el valor constante.

$$\overline{X}_n = \frac{(X_1K)+(X_2K)+(X_3K)+(X_4K)+\cdots(X_nK)}{n}$$

$$= \frac{(X_1 + X_2 + X_3 + X_4 + \cdots X_n)K}{n}$$

$$= \frac{K\sum_{i=1}^{n} X_i}{n}$$

$$= K\ \overline{X}$$

Ejemplo:

El peso de 100 estudiantes es el siguiente; si los pesos aumentan un 5%, ¿cuál será el nuevo promedio?

Tabla 12.
Distribución de los peso de los estudiantes

Pesos (Kg),	No, de estudiantes		Pesos aumentado el 5% (Kg)	
x_i	f_i	$x_i f_i$	x_{i2}	$x_{i2}\ f_i$
45,5	6	273	47,8	286,7
48,3	8	386,4	50,7	405,7
50,5	10	505	53,0	530,3
55,5	13	721,5	58,3	757,6
68,6	22	1509,2	72,0	1584,7
68,7	12	824,4	72,1	865,6
71,4	9	642,6	75,0	674,7
72,5	8	580	76,1	609,0
75,5	7	528,5	79,3	554,9
78,5	5	392,5	82,4	412,1
	100	6363,1		6.681,3

Fuente: Elaboración propia.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i f_i}{n} = \frac{6363,1}{100} = 63,631$$

$$\overline{X}_n = \frac{6.681,3}{100} = 66.813$$

Aplicando la propiedad obtenemos

$$\overline{\mathrm{X}} = (63.631)\ 0.05 + 63.631 = 66.813$$

En la práctica debe utilizarse la propiedad.

Quinta: Si los diferentes *n* datos de una variable se descomponen en muestras de $n_1, n_2, n_3 \ldots.. n_k$ elementos, y cada posee su propio promedio $X_1, X_2, X_3 \ldots\ldots X_k$ entonces el promedio general de todos los n datos será igual al promedio de los varios promedio de la muestra. Esto es:

$$X_G = \frac{X_1(n_1) + X_2(n_2) + X_3(n_3) + \ldots\ldots X_k(n_k)}{n_1, n_2, n_3 \ldots.. n_k}$$

$$X_G = \frac{\sum_{i=1}^{n} X_i(n_i)}{\sum_{i=1}^{n} n_i} = \frac{\sum_{i=1}^{n} X_i(n_i)}{n}$$

Esta propiedad expresa la forma de hallar el promedio ponderado general entre varias muestras, tomando como ponderaciones los tamaños de cada muestra.

Ejemplo:

Si tienen 4 grupos de estudiantes cursando una misma materia; por ejemplo matemáticas, y se disponen de los siguientes datos:

Tabla 13.
Grupos de estudiantes cursando una misma materia

Grupos	Promedio obtenido (x_i)	No. de estudiantes (n_i)	(x_i) (n_i)
A	3.5	25	87.5
B	4.2	15	63
C	2.8	40	112
D	3.0	10	30
	TOTAL	90	292.5

Fuente: Elaboración propia.

Aplicando la quinta propiedad, la forma de obtener el promedio general de los 90 estudiantes será:

$$X_G = \frac{\sum_{i=1}^{n} X_i (n_i)}{n} = \frac{292.5}{90} = 3.25$$

El promedio general de todos los estudiantes es de 3.25

Debe observarse, que la tendencia general de todos los estudiantes es sumar los promedios de los 4 grupos y dividir por el número de grupo, lo cual es erróneo. *Esto debe corregirse.*

$$X = \frac{3.5 + 4.2 + 2.8 + 3.0}{4} = \frac{13.5}{4} = 3.375$$

Lo cual es completamente diferente al verdadero promedio general que fue 3.25.

La Mediana

Si en un conjunto de valores observamos que la tendencia de los datos esta sesgada hacia los valores altos o hacia los valores bajos entonces es inconveniente utilizar como valor representativo de estos datos el promedio Debido a que este estaría influenciado por estos valores extremos. Entonces sería más conveniente utilizar un indicador que nos refleje la situación planteada o que fuese más representativo este indicador es la *mediana* o *valor medio*, y se simboliza por las letras “*Me*”.

Definición

La mediana está dada por el valor ubicado en el centro de todos los datos cuando éstos estén ordenados del mínimo al máximo. La mediana se representa por el símbolo *Me*.

Para encontrar la mediana, primero se ordenan los datos de una muestra de menor a mayor y se ubica el que está en el medio. En el caso en que la cantidad de datos sea impar, la mediana es el valor que está en el centro. Si la cantidad de dato es par, se saca el promedio de los dos datos que se encuentran ubicados en el centro de la distribución.

Forma de cálculo

Existen 2 maneras de calcular la mediana de los datos de una variable.

- *La mediana para datos no agrupados*

Si se dispone de pocos datos ($n \leq 30$) se procede de la siguiente manera:

a) Si el número de datos es impar, y los datos han sido ordenados en orden creciente (o decreciente), $X_1, X_2, X_3 \ldots. X_n$, entonces la mediana será aquel valor que este situado en la mitad de los datos así ordenados.

Esta es la mediana cuando es igual a:

$$Me = X_{\frac{(n+1)}{2}}$$

b) Si el número de datos es par, y se ha procedido a su ordenamiento, entonces la mediana será igual al promedio de los 2 valores centrales:

$$Me = \frac{X_{(n/2)} + X_{(n/2+1)}}{2}$$

Ejemplo:

Sea los valores 21, 8, 1, 3, 4, 1, 9, 9, 9, 2, 15, 2, 7 entonces n = 13 (impar).

Ordenamos los valores de la menor a la mayor magnitud.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
1	1	2	2	3	4	5	5	7	8	9	15	21

$$Me = X_{\frac{(n+1)}{2}} = X_{\frac{(13+1)}{2}} = X_{\frac{(14)}{2}} = X_7 = 5$$

Es decir, la mediana corresponde al dato ubicado en la séptima posición que es igual a 5.

Ejemplo:

Sea los valores 1230, 7, 420, 2, 8, 3, 24, 75, 12, 6, 9, 29, 68, 39, 45, 6, 20, 52.

Entonces, n = número de datos = 18 datos (par)

Ordenados estos valores de la menor a la mayor magnitud:

3, 6, 6, 7, 8, 9, 12, 20, 24, 29, 39, 45, 52, 68, 72, 75, 420, 1230

$$Mediana = Me = \frac{X_{(n/2)} + X_{(n/2+1)}}{2} = \frac{X_{(18/2)} + X_{(18/2+1)}}{2} = \frac{X_9 + X_{10}}{2}$$

$$= \frac{24+29}{2} = \frac{53}{2} = 26.5$$

- *La mediana para datos agrupados*

Si, por el contrario, se tiene una gran cantidad de datos, lo cual hace necesario que se ordenen en una tabla de distribución de frecuencia, o cuando la información suministrada de los datos se encuentra en una tabla de distribución de frecuencia. Para calcular la Mediana se realizan los siguientes pasos:

- Primero se encuentra el valor n / 2
- Segundo se identifica el intervalo de clase donde está la media. Para encontrarlo se ubica el intervalo de clase donde la frecuencia acumulada sea mayor o igual al valor de n/2.
- Tercero hacemos uso de la siguiente fórmula de la mediana:

$$Me = L_i + I\left[\frac{\frac{n}{2} - F_{i(Me)}}{f_{i(Me)}}\right]$$

Donde:

n = Numero de datos de distribución o muestra
L_i = Límite inferior intervalo o clase mediana.
I = Intervalo o amplitud de cada clase.
$F'_{(Me)}$ = frecuencia absoluta acumulada que corresponde a la clase anterior a la clase o intervalo mediano.
$f_{i(Me)}$ = frecuencia absoluta correspondiente al intervalo mediano.

Utilizando el modelo apropiado, calculemos el valor de la mediana:

Tabla 14.
Distribución de Frecuencias Correspondiente a los Datos de las Estaturas de 50 Estudiantes

Intervalo	Fi	X_i	$h_{i\,(\%)}$	F_i	H_i (%)	(x_i) (f_i)	$X_i\,f_i$
50.5-53.5	1	52	0.02	1	0.02	52	1.04
53.5-56.5	2	55	0.04	3	0.06	110	2.20
56.5-59.5	6	58	0.12	9	0.18	348	6.96
59.5-62.5	11	61	0.22	20	0.4	671	13.42
62.5-65.5	16	64	0.32	36	0.72	1024	20.48
65.5-68.5	9	57	0.18	45	0.9	523	10.26
68.5-71.5	4	70	0.08	49	0.98	280	5.60
71.5-74.5	1	73	0.02	50	1.00	73	1.46
Total	50		1.00			3.71	61.42

Fuente: Elaboración propia.

Ahora, si tan solo se tuviera la información de frecuencias en forma de porcentajes (*hi*) y se deseara encontrar el valor de la mediana se procedería de la siguiente manera, con el modelo que se da a continuación:

$$Me = L_i + I\left[\frac{\frac{1}{2} - H_{i(Me)}}{h_{i(Me)}}\right] = 62.5 + 3\left[\frac{0.5 - 0.4}{0.32}\right]$$

$$= 62.5 + 3\left[\frac{0.1}{0.32}\right] = 62.5 + 0.9375 = 63.4375 = 63.44$$

Se observa que este resultado es el mismo que el obtenido utilizando el procedimiento anterior. Al hacer los cálculos de la mediana, esta no se encuentra afectada por los valores extremos de la distribución.

Lo que interesa, al calcular la mediana, es la posición central que ocupa los valores, y no su magnitud, alta o baja. Por el contrario, el promedio si es afectado por los valores extremos de la distribución, puesto que es necesario sumar sus magnitudes y los valores extremos o altos (o bajos) afectan por consiguiente esta suma.

Ejemplo:

Sean los siguientes datos referentes a las edades de unas 10 personas:

$$25, 84, 37, 26, 15, 3, 45, 64, 17, 31$$

$$\text{Pr}\,omedio = \frac{25+84+37+26+15+3+45+64+17+31}{10} = \frac{347}{10} = 34.7$$

Mediana ordenando los datos obtenidos.

$$3, 15, 17, 25, 26, 31, 37, 45, 64, 84$$

$$Me = \frac{26+31}{2} = 28.5$$

Cuartiles, Deciles y Percentiles

Otras medidas que estan muy asociadas a la mediana que se basan también en su posición en una serie de observaciones, son los cuartiles, deciles y percentiles.

Si las series de observaciones se dividen en cuatro partes iguales los valores correspondientes de la serie toma el nombre de CUARTIL, denotados como Q_1, Q_2, Q_3.

> *El primer cuartel Q_1 es aquel valor que tiene por debajo el 25% de los valores de los datos; el tercer cuartel Q_3 es el valor que tiene por debajo el 75% de los datos. Y el segundo cuartel, Q_2 es exactamente la mediana.*

Para datos sin agrupar el cuartil, al igual que la mediana, toma uno de los valores de la variable, o el valor intermedio entre dos valores. Si n/4 es entero, el primer cuartil toma el valor situado entre los números n/4 y el valor mayor que le siga. Si n/4 no es entero, el primer cuartil toma el valor del mayor valor siguiente. Si se sustituye n/4 por 3n/4, se puede de igual manera encontrar el valor del tercer cuartil.

Para datos agrupados en tablas, por ejemplo, el Q_1, será 61 y Q_3 será 66.

Para datos agrupados en tabla de distribución de frecuencia el método de cálculo de los cuartiles es parecido al de la mediana. El cálculo de K – décimo cuartil está dado por:

$$Q_K = L_i + I\left[\frac{\frac{Kn}{4} - F_{i(K)}}{f_{i(K)}}\right]$$

Ejemplo:

Utilizando los datos anteriores de la tabla tenemos:

$$Q_1 = L_i + I\left[\frac{\frac{n}{4} - F_{i(1)}}{f_{i(1)}}\right] = 59.5 + 3\left[\frac{\frac{50}{4} - 9}{11}\right] = 59.5 + 3\left[\frac{3.5}{11}\right] = 60.49$$

$$Q_3 = L_i + I\left[\frac{\frac{3n}{4} - F_{i(3)}}{f_{i(3)}}\right] = 65.5 + 3\left[\frac{\frac{3(50)}{4} - 36}{9}\right] = 65.5 + 3\left[\frac{1.5}{9}\right] = 66$$

Los percentiles dividen la distribución en 100 partes iguales, y se calculan mediante la misma técnica utilizada para la mediana y los cuartiles, con ligeras variaciones. Para datos no agrupados, el valor percentil forma el valor intermedio entre dos observaciones, según que n sea o no divisible por 100. Si los datos están agrupados se utiliza la siguiente relación, parecida al de la mediana y los cuartiles:

$$P_K = L_i + I\left[\frac{\frac{Kn}{100} - F_{i(K)}}{f_{i(K)}}\right]$$

Ejemplo:

$$P_{60} = 62.5 + 3\left[\frac{\frac{60(50)}{100} - 20}{16}\right] = 62.5 + 3\left[\frac{10}{16}\right] = 62{,}5 + 1{,}875 = 64{,}375$$

De igual manera, se hace el cálculo de los deciles, los cuales dividen la distribución de frecuencia de una variable en dos partes iguales.

La Moda

Con mucha frecuencia se clasifican los datos en grupos numéricos, es decir cuando se agrupan *por una cualidad especifica.* Como por ejemplo, los empleados de una compañía se pueden clasificar por sexo, estado civil, ocupaciones, edad, etc. En tales circunstancias no tiene sentido referirse a la media o mediana del sexo de los empleados, de su estado civil, o de sus ocupaciones, hasta tanto no se asignen valores cuantitativos a los diversos grupos de esta variable.

Pero si tendrá sentido referirse por e*jemplo,* al formular la pregunta de cuál será la ocupación de la mayoría de los empleados. Este dato de ocupación se denomina entonces la *ocupación modal.* Por lo tanto se hace necesario introducir este nuevo concepto de medida central, llamado *moda* o *valor Modal* cuyo símbolo es *Mo*.

La *moda* se define como el valor que ocurre con mayor frecuencia en una distribución.

Para hallar el valor modal, es necesario localizar la clase modal en la cual se sitúa. En general, la clase modal es aquella que tiene la máxima frecuencia de la distribución. Una vez hallada la clase modal, el paso siguiente es localizarla dentro de esa clase; el punto de la clase modal sería entonces el valor de la distribución si tuviera un número infinito de datos de los intervalos de clase y fueran ínfimamente pequeños. La razón para esto es que el intervalo de clase se reduce al valor modal que se encuentra dentro la amplitud menor de valores, y así tienden a acercarse al punto medio de la clase. Pero en la práctica esta situación ideal no existe; por consiguiente el valor modal se estima por el método de interpolación, utilizando para ello la siguiente relación:

$$Mo = L_i + I\left[\frac{\Delta_1}{\Delta_1 \quad \Delta_2}\right]$$

Donde:

Δ_1 = diferencia absoluta entre la frecuencia modal y la frecuencia anterior.
Δ_2 = diferencia absoluta entre la frecuencia modal y la frecuencia siguiente.
I = intervalo de clase.
L_i = límite inferior de la clase modal.

Si aplicamos esta fórmula a los datos de la tabla 14 tendremos.

$$Mo = 62.5 + 3\left[\frac{5}{5+7}\right] = 62.5 + \left[\frac{15}{12}\right] = 63.75$$

Estimar la moda tiene varios inconvenientes:

Primero, el estimado en una distribución asimétrica esta demasiado cercano a un extremo de los datos y no puede ser un buen representante de la distribución.

Segundo, la localización de la clase modal depende de la manera como se hallan clasificados los datos.

Tercero, en cierto caso puede ser muy vaga en su utilidad.

2. MEDIDAS DE VARIABILIDAD

Con frecuencia se analizan datos o resultados de series de datos de la misma variable, con base en los resultados del promedio. Así, en dos conjuntos de datos comparados, se puede llegar a conclusiones erróneas, si sólo se analizan sus promedios.

Ejemplo 1

Se desea elegir un empleado de tres aspirantes para trabajar en el control de calidad de los televisores que produce la empresa XZ. Para su selección se realiza una prueba a los aspirantes sobre diez tópicos que contemplan la capacidad intelectual, destreza y relación con los demás. ¿A cuál persona se debe escoger con base en el mejor puntaje y desempeño ante la prueba realizada? Los resultados obtenidos por cada uno de ellos fueron:

Tabla 15.
Puntaje de los candidatos

CANDIDATOS	A	B	C
PUNTAJE	65	95	68
	35	65	57
	96	86	55
	58	78	68
	24	36	59
	88	58	68
	36	70	53
	98	93	52
	46	20	65
	75	25	75
TOTAL	621	626	620
PROMEDIO	62.1	62.6	62.0

Fuente: Elaboración propia.

Solución:

Se observa que los puntajes promedios son muy similares; y a duras penas podríamos decir sí el candidato B fue el que obtuvo el mayor promedio. Pero este análisis es incompleto, puesto que hay que tener en cuenta la homogeneidad de los resultados o sea en la variabilidad de cada candidato en sus puntajes.

En otras palabras, es necesario incorporar un elemento más de análisis en el comportamiento de los resultados obtenidos por los tres candidatos. Es necesario analizar qué tan variables son los puntajes de los candidatos con respeto a su a promedio.

Entonces es lógico suponer que el candidato que tiene menos variabilidad en los puntajes obtenidos en las pruebas, es el candidato más estable y más constante en sus puntajes. Esta forma de medir la variabilidad de los resultados con relación a un valor central o promedio, se denomina *media de dispersión* o *de variabilidad.*

Si graficamos los resultados obtenidos por el candidato C, es el que menos se aleja al promedio general de los 3 grupos (X = 62.2 puntos). Con este nuevo análisis, se puede concluir que el candidato para escoger, es el candidato C, por tener menos variabilidad con sus puntajes.

Ahora que se puede calcular estas medidas de variabilidad; los principales parámetros de variación que se van a analizar son en su orden: la *varianza, desviación Standard, y con coeficiente de variabilidad.*

La Varianza de una Población

La varianza representada como σ^2 *(sigma cuadrado)* es una medida de dispersión, que se define como el promedio del cuadrado de la desviación de dicha variable respecto a su media aritmética. Es decir.

$$Varianza = \sigma^2 = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + (X_3 - \mu)^2 + \cdots (X_n - \mu)^2}{N}$$

$$= \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

Esta fórmula, se utiliza cuando los datos no estan agrupados, es decir, no han sido tabulados en tablas de distribución de frecuencias.

Si el número de datos de la población es bastante grande, se hace necesario que los datos se resuman en una tabla de frecuencia, entonces la varianza poblacional se calcula de la siguiente manera:

$$Varianza = \sigma^2 = \frac{(X_1-\mu)^2 f_1 + (X_2-\mu)^2 f_2 + (X_3-\mu)^2 f_3 + \cdots (X_K-\mu)^2 f_K}{f_1 + f_2 + f_3 + \cdots + f_K}$$

$$= \frac{\sum_{i=1}^{N} (X_i-\mu)^2 f_i}{N}$$

Forma de cálculo

Al calcular la varianza, se debe observar lo siguiente:

1. En primer lugar, es necesario calcular el promedio "μ".
2. Para cada valor X se calcula la desviación con respecto al promedio.
3. Cada desviación obtenida, ya sea positiva o negativa se eleva al cuadrado.
4. Cada desviación elevada al cuadrado, se multiplica por su correspondientes frecuencias " f_i ".
5. Se suman todas estas cantidades de desviaciones al cuadrado.
6. Y por último, se dividen estas cantidades obtenidas por el número de datos utilizado (N).

Ejemplo 2

Se tienen los siguientes datos correspondientes a una variable:

Tabla 16.
Cálculo de varianza

VALOR X_I	$(X_I - \mu)$	$(X_I - \mu)^2$
25	-30	900
37	-18	324
45	-10	100
58	+ 3	9
65	+10	100
80	+25	625
45	-10	100
65	+10	100
90	+35	1.225
40	-15	225
550	0	3.708

Fuente: Elaboración propia.

$$\mu = \frac{550}{10} = 55$$

$$Varianza = \sigma^2 = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{N} = \frac{3708}{10} = 370{,}8$$

Ahora, si los datos se encuentran distribuidos en una tabla de frecuencia, su cálculo será como el ejemplo siguiente:

Ejemplo 3

Temperatura. °C

Tabla 17.
Distribución de la temperatura

Valor X_I (Intervalos) °c	Frecuencias	Valor medio (X_i)	(X_i) (f_i)	$(X_i - \mu)^2$	$(X_i - \mu)^2$
-8 - -2	15	-3	-75	337.0896	5056.334
-2 – 4	35	1	35	152.7696	5346.936
4 – 10	55	7	385	40.4496	224.728
10 – 16	68	13	884	0.1296	8.812
16 – 22	52	19	988	31.4896	1654.099
22 – 28	30	25	750	135.4896	4064.688
28 – 34	25	31	775	311.1696	7779.240
	280		3.742		26.134.837

Fuente: Elaboración propia.

$$\mu = \frac{3742}{280} = 13.36$$

$$Varianza = \sigma^2 = \frac{\sum_{i=1}^{n} (X_i - \mu)^2 f_i}{N} = \frac{26134837}{280} = 93{,}33$$

La varianza, se expresa en unidades tomadas al cuadrado y por lo tanto no expresa la misma cantidad de la variable original, debido a la elevación al cuadrado. Esto es inconveniente, puesto que este parámetro no permite analizar los resultados en forma conveniente.

Entonces es necesario volver a expresar el parámetro en las unidades originales de variable; por eso extraemos de la raíz cuadrada a esta expresión y obtenemos otro nuevo parámetro de variabilidad llamado *división Standard*.

Desviación Estándar

Definición

Se simboliza por σ (*sigma*). La *desviación estándar* es una medida de variabilidad que indica la dispersión promedio de una variable, con respecto a su promedio. Es la raíz cuadrada de la varianza.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu)^2 f_i}{N}} = \sqrt{Varianza}$$

En nuestro ejemplo 2, calculemos la desviación estándar, así:

$$\sigma = \sqrt{Varianza} = \sqrt{370.8} = 19.25$$

Esto indica, que los datos X_i se desvían en promedio19.25 unidades de su valor medio $\mu = 55$.

Para el ejemplo 3, la desviación estándar será:

$$\sigma = \sqrt{Varianza} = \sqrt{93.33} = 9.66$$

Así se puede afirmar que la temperatura tiene una desviación promedia de 9.66 con relación a su valor promedio 13.36 grados.

Si en el ejemplo 1 se analizan los 3 candidatos con base en sus puntajes promedio y los valores de su desviación estándar, se debe entonces escoger como mejor candidato, aquel que tenga menor división estándar.

Entonces, al calcular sus valores:

Candidato A:

$$Varianza = \sigma^2 = \frac{(65-62.1)^2 + (35-62.1)^2 + (96-62.1)^2 + \cdots + (75-62.1)^2}{10}$$

$$= \frac{8.41 + 734.41 + 1149.21 + 166.41}{10}$$

$$= \frac{6426.9}{10} = 442.69$$

$$Desviación \quad es\tan dar = \sigma = \sqrt{442.69} = 25.35$$

Candidato B:

$$Varianza = \sigma^2 = \frac{(95-62.6)^2 + (65-62.6)^2 + \cdots + (26-62.6)^2}{10}$$

$$= \frac{1049.76 + 5.76 + 547.56 + \cdots + 1413.76}{10}$$

$$= \frac{6776.4}{10} = 677.64$$

$$Desviación \quad es\tan dar = \sigma = \sqrt{677.64} = 26.03$$

Candidato C

$$Varianza = \sigma^2 = \frac{(68-62)^2 + (57-62)^2 + (55-62)^2 + \cdots + (75-62)^2}{10}$$

$$= \frac{36 + 25 + 49 + \cdots + 69}{10}$$

$$= \frac{550}{10} = 55$$

$$Desviación \quad es \tan dar = \sigma = \sqrt{55} = 7.41$$

Tabla 18.
Promedios y desviación estándar de los candidatos

	promedios	Desviación Standard
Candidato A	62.1	25.35
Candidato B	62.6	26.03
Candidato C	62.0	7.41

Fuente: Elaboración propia.

Segun los resultados obtenidos, tenemos que el candidato con menos desviación del valor promedio, es C. por consiguiente es el que se debe seleccionar.

Desviación Estándar de la Muestra

Cuando se calcula la varianza y la desviación estándar de una muestra, se procede de la misma forma como se explicó anteriormente, pero reemplazando el valor del promedio poblacional "*u*", por su correspondiente valor muestral *X*, y en vez de dividir por *N* (tamaño de la población) se divide por $n - 1$ (n = Tamaño Muestral).

Así, cuando se desea calcular la varianza y la desviación estándar de una muestra, se procede a utilizar las siguientes formulas.

Varianza muestral:

$$S^2 = \frac{\sum_{i=1}^{n} \left(X_i - X_2\right)^2 f_i}{n-1}$$

Desviación muestral

$$S = \sqrt{\frac{\sum_{i=1}^{n} \left(X_i - X_2\right)^2 f_i}{n-1}}$$

Ejemplo:

Si se desea encontrar la desviación estándar de una muestra tomada aleatoriamente a una muestra de 8 datos, se procede de la siguiente manera:

Datos: (peso Kg.) X_i = 12, 55, 17, 28, 13, 25, 34, 56,

Calculamos el peso promedio:

$$X = \frac{12+55+17+28+13+25+34+56}{8} = \frac{240}{8} = 30$$

Desviaciones de los datos con respecto al promedio:

Varianza muestral

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - X_2)^2}{n-1} = \frac{2128}{8-1} = \frac{2128}{7} = 304$$

Desviación muestral

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - X_2)^2}{n-1}} = \sqrt{304} = 17.43$$

Existe una fórmula más práctica para calcular la varianza y la desviación Estándar de datos tabulados en tablas de distribución de frecuencia:

Varianza muestal

$$S_x^2 = \frac{n\sum_{i=1}^{n} X_i^{\,2} - \sum_{i=1}^{n}(X_i)^2}{n\ (n-1)}$$

Desviación Estándar muestral

$$S_x = \sqrt{\frac{n\sum_{i=1}^{n} X_i^{\,2} - \sum_{i=1}^{n}(X_i)^2}{n\ (n-1)}}$$

$$S_x = \sqrt{\frac{8(9328)-(240)^2}{8\ (8-1)}} = \sqrt{\frac{74624-57600}{56}}$$
$$= \sqrt{\frac{17024}{56}} = \sqrt{304} = 17.43$$

Es el mismo resultado obtenido con anterioridad, lo cual indica que las formulas dan idénticos resultados.

Propiedades de la Varianza y Desviación Estándar

- **Primera Propiedad:** La varianza de una variable, y por consiguiente la desviación estándar será siempre una cantidad positiva, es decir igual o mayor a cero.

Esto es, $\sigma^2 \geq 0$; $\sigma \geq 0$ (población)

$S^2 \geq 0$; $S^2 \geq 0$ (muestra)

Si la varianza, o la desviación Standard, es exactamente igual 0, entonces los valores de la variable son todos iguales a una constante K

- **Segunda Propiedad:** Si todos los datos de una muestra son iguales, entonces la desviación es cero. Así:

$$\sigma^2 = \frac{(K_1-K)^2+(K_2-K)^2+\cdots+(K_n-K)^2}{n}$$
$$\sigma^2_{(k)} = \frac{0}{n} = 0$$

- **Tercera Propiedad**: La varianza de los datos de una variable a las cuales se le suma (o se resta) un valor constante, será igual a la misma varianza original.

Esto significa que la varianza es invariante, si se le suma (o se resta) un mismo valor constante.

Tabla 19.
Distribución de los datos

X_I	f_i	$X_I f_i$	$(X_I - X)^2 f_i$	$X_I + K$	f_i	$(X_I + K - {}_n)^2 f_i$
10	2	20	1.372.88	15	2	1.372.88
25	6	150	752.64	30	6	752.64
28	10	280	672.40	33	10	672.40
35	4	490	20.16	40	14	20.16
45	6	270	5.664.40	50	6	464.64
60	10	600	5.664.40	65	10	5.664.40
	50	1810	8.947.12			8.947.12

Fuente: Elaboración propia.

Ejemplo:

Sean los datos siguientes.

* aplicando la propiedad

$$Promedio \quad X = \frac{1810}{50} = 36.2$$

$$Varianza \quad S_x^2 = \frac{8947.12}{49} = 36.2$$

Sea $K = 5$

* *Nuevo promedio* aplicando la 3ª propiedad de la media.

Media de la variable más la constante = $X + K$.

Nuevo promedio $X_n = 36.2 + 5 = 41.2$

* Nueva Varianza:

$$S^2\ (X + K) = S^2{}_{(X)} + S^2{}_{(K)} = S^2(X) + 0 = S^2{}_{(X)}$$

O sea que la nueva varianza *O* sea una nueva varianza a la misma:

$$S^2\ (X + K) = 182.59 + 0 = 182.59$$

En la práctica, se aplica directamente la propiedad, y no se hace el cambio numérico de los valores de la variable.

- **Cuarta Propiedad:** Si el valor de una variable se multiplica por un mismo valor, entonces la nueva varianza será igual a la varianza de la variable multiplicada por el cuadrado de las constantes *K*.

Esto es la varianza, la varianza será igual a:

$$\sigma^2 = [K.\ X_i] = K.\ \sigma^2{}_{(x)}$$

La desviación estándar = $\sigma^2 = [K.\ X_i] = K.\ [\sigma_{(xi)}]$

Ejemplo:

Sean los valores X_i:

3, 7, 16, 20, 32, 45, 50, 65, 70, 92

$$\text{Promedio } X = \frac{400}{10} = 40$$

$$Varianza \quad S_x^2 = \frac{7852}{9} = 872.44$$

Sea $K = 3$

Nuevos valores = ($X_1 . K$) =

X_1 = 3, 7, 16, 20, 32, 45, 50, 65, 70, 92

($K. X_1$) = 9, 21, 48, 60, 96, 135, 150, 195, 210, 276

$$Varianza = \frac{\sum_{i=1}^{10} \left[(X_i - K) - X \right]^2}{10 - 1} = \frac{70668}{9} = 7852$$

Aplicando la propiedad, tendremos:

Nueva varianza:

$$S^2_{(x\,.\,k)} = K^2 \,.\, S^2_X = 3^2\,(872.44) = 7.852$$

Obsérvese que aplicando la propiedad, se obtiene el mismo resultado. En la práctica, no se hacen todos estos cálculos, sólo se aplica directamente la propiedad.

Ejemplo:

Tres máquinas diferentes de una fábrica de muebles se utilizan para cortar trozos de madera para una cierta pieza de un mueble. El ajuste en la máquina es parte importante en el funcionamiento adecuado de las máquinas. La máquina No. 1 tiene un corte promedio de 2 mts., con una desviación estándar de 0.005 mts.; la máquina No. 2 tiene un corte promedio de 3 mts. con una desviación estándar de 0.07 mts., y la máquina No. 3 tiene un promedio de 4 mts. con una desviación de 0.1 mts.

Se seleccionan 3 piezas de madera, una de cada máquina y se les mide su longitud. La pieza de la maquina No. 1 midió 1,994 Mts; de la máquina No. 2 midió 3.007 Mts, y la pieza de la máquina No. 3 midió 3,9907 Mts.

¿Cuál de las 3 piezas de madera está mal cortada en relación a su propia máquina?

Solución:

Tabla 20.
Distribución de las piezas por maquina

Maquina	Promedio	Desviación Standard	Longitud de las Piezas	Desviación
No. 1	$X_1 = 2$ Mts	$S_1 = 0.005$	1.994 Mts	1.994 – 2 = 0.006 mts
No. 2	$X_2 = 3$ Mts	$S_2 = 0.007$	3.007 Mts	3.007 – 3 = 0.007 mts
No. 3	$X_3 = 4$ Mts	$S_3 = 0.1$	3.990 Mts	3.990 – 4 = 0.01 Mts

Fuente: Elaboración propia.

Al analizar las desviaciones de cada trozo de madera, se observa que la pieza que más se desvía de su respectivo valor es la maquina número 1. Por lo tanto, esta pieza es la defectuosa o está mal cortada.

Coeficiente de Dispersión o Variación "Δ" (C.V.)

El coeficiente de variación está dado por el cociente obtenido de dividir la desviación estándar entre la media aritmética. Este coeficiente también puede representarse en forma porcentual y es importante resaltar que los valores sean positivos. Si el valor del coeficiente de variación es grande, se relaciona con una mayor heterogeneidad de los datos; y cuando el coeficiente de variación se pequeño, se relaciona con una mayor homogeneidad de los datos.

Se calcula:

$$Cv = \delta = \frac{\sigma}{\left|\bar{x}\right|}$$

Donde σ es la desviación típica, y $\bar{x}$ es la Media.

Se puede dar en porcentaje calculando:

$$Cv = \delta = \frac{\sigma}{\left|\bar{x}\right|}100\%$$

Ejemplo:

Un operador A produce 40 unidades por hora en un proceso industrial, con la desviación estándar de 5. Otro operador produce 160 unidades por hora con una desviación estándar de 15. ¿Cuál de las dos operaciones muestra una menor variabilidad?

Solución:

A simple vista parece ser que el operador B tiene 3 veces más variabilidad que el operador A pero debe tenerse en cuenta que el operador B produce unidades a una taza 4 veces mayor que el operador A.

Entonces, con el objeto de analizar mejor esta información estadística de los dos operarios, comparemos sus correspondientes valores de coeficiente de variación.

$$\delta_A = \frac{\sigma_A}{\mu_A}\,100 = \frac{5}{4}\,100 = 12.5\%$$

$$\delta_B = \frac{\sigma_B}{\mu_B}\,100 = \frac{15}{160}\,100 = 9.4\%$$

De esta manera, el operador B tiene mayor variabilidad absoluta (σ = 15), y una menor variación relativa (σ = 9.4%) por que la medida de producción del operador D es mucho mayor que la del operador A.

3. EJERCICIOS RESUELTOS

Ejemplo 1

En la tabla siguiente se registra el número de empleados de las 60 empresas más importantes de Colombia.

80	86	91	96	101	107	111	81	88	83
99	103	106	112	82	89	95	98	105	108
113	83	94	97	102	109	114	84	92	98
104	107	115	80	93	97	108	112	84	94
96	107	114	83	99	106	113	82	97	112
81	96	111	98	112	98	113	114	112	113

Con base en esta información calcular los siguientes tópicos de estadística:

1. Una distribución de frecuencia.
2. La construcción de los histogramas de frecuencia absoluta, acumulada y ojiva
3. Calcular la media aritmética, mediana, la moda y el primer cuartil.
4. Hacer el cálculo y la interpretación de la varianza y desviación estándar.

Solución:

1. Una distribución de frecuencia con 7 intervalos de clase

Primero se organiza la tabla en forma ascendente

80	80	81	81	82	82	83	83	83	84
84	86	88	89	91	92	93	94	94	95
96	96	96	97	97	97	98	98	98	98
99	99	101	102	103	104	105	106	106	107
107	107	108	108	109	111	111	112	112	112
112	112	113	113	113	113	114	114	114	115

Cálculo del Rango:
Rango = Límite Superior - Límite Inferior =
115 –80 = 35

Número de intervalos:
Para obtener un valor aproximado, podemos emplear la regla de ¨*Sturges*¨.

$$k = 1 + 3.3 * \log(n) = 1 + 3.3 * \log(60) = 6.86789913 \cong 7$$

Luego el número de intervalos es 7.

Tamaño del Intervalo:
RANGO / NÚMERO DE INTERVALOS =
35/7 = 5

Tabla 21.
Distribución del número de empleados

INTERVALOS		ni	Ni	hi	Hi	Xi	Xi * ni	|Xi-M| ni	|Xi-Me| ni	Xi-M	(Xi - M)^2	Ni (Xi - M)^2
80	85	11	11	18,33%	18,33%	82,5	907,5	184,8	194,94444	-16,8	282,24	3104,64
86	91	4	15	6,67%	25,00%	88,5	354	43,2	46,888889	-10,8	116,64	466,56
92	97	11	26	18,33%	43,33%	94,5	1039,5	52,8	62,944444	-4,8	23,04	253,44
98	103	9	35	15,00%	58,33%	100,5	904,5	10,8	2,5	1,2	1,44	12,96
104	109	10	45	16,67%	75,00%	106,5	1065	72	62,777778	7,2	51,84	518,4
110	115	15	60	25,00%	100,00%	112,5	1687,5	198	184,16667	13,2	174,24	2613,6
		60		100,00%		585	5958	561,6	554,22222	-10.8	649.44	6969,6

Fuente: Elaboración propia.

2. La construcción de los histogramas de frecuencia absoluta, acumulada y ojiva

A. Histogramas

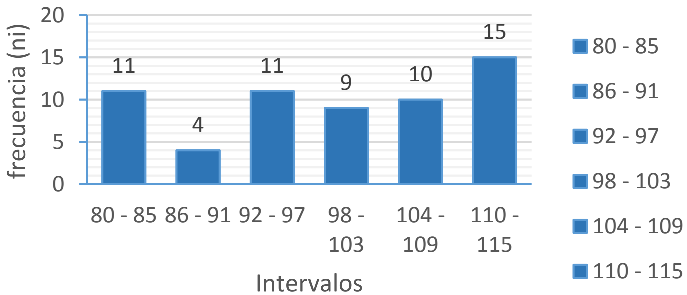

Figura 7. ***Distribución de frecuencia relativa***
Fuente: Elaboración propia.

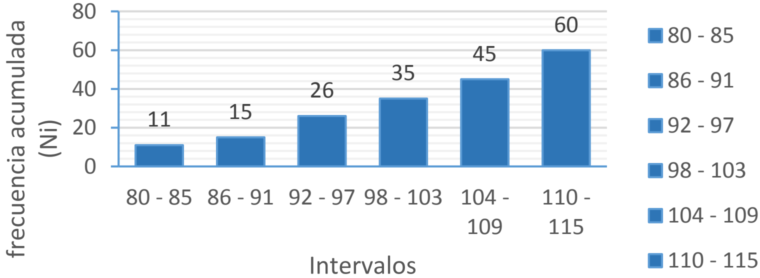

Figura 8. ***Distribución de frecuencia acumulada***
Fuente: Elaboración propia.

B. Polígono

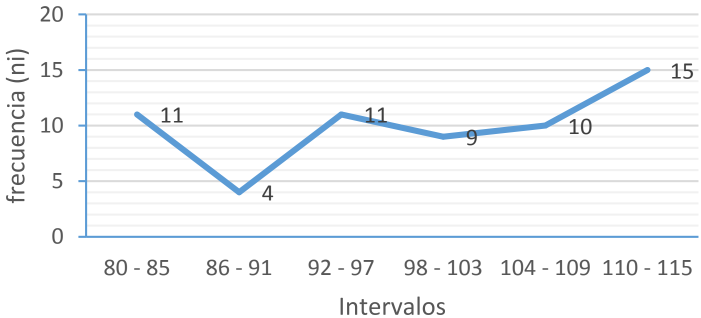

Figura 9. ***Polígono de frecuencia***
Fuente: Elaboración propia.

C. Ojiva

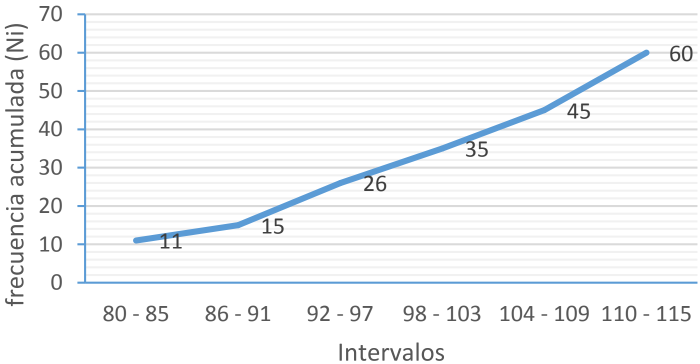

Figura 10. ***Ojiva***
Fuente: Elaboración propia.

3. Calculo de la media aritmética, mediana, la moda y el primer cuartil. (*estadígrafos*).

a) La media aritmética:

$$\text{M} = \text{MEDIA ARITMÉTICA} = \frac{\sum X_i n_i}{n} = \frac{5958}{60} = 99{,}3$$

b) La mediana

$$\text{Me} = \text{MEDIANA} = \quad Z_i + a\left(\frac{\frac{n}{2} - N_{i-1}}{N_i - N_{i-1}}\right)$$

Con:

Zi = 98	a = 5	n/2= 60/2 = 30	Ni = 35	Ni-1 = 26

Me = MEDIANA = $98+5\left(\dfrac{30-26}{35-26}\right)=100{,}22$

c) La Moda

Mo = MODA = $Z_i+a\left(\dfrac{n_i-n_{i-1}}{2n_i-n_{i-1}-n_{i+1}}\right)$

Con:

Zi = 110	a = 5	ni = 15	ni-1 = 10	ni+1 = 0

Mo = MODA = $110+5\left(\dfrac{15-10}{2\,(15)-10-0}\right)=111{,}25$

d) Cuartiles

Primer cuartil

$$Q_{\frac{1}{4}}=Li+a\left[\dfrac{\dfrac{N}{4}-N_{i-1}}{N_i-N_{i-1}}\right]\Rightarrow\ \dfrac{N}{4}=\ \dfrac{60}{4}\ =15\Rightarrow N_i\geq 15$$

$$Q_{\frac{1}{4}}=86+5.\left[\dfrac{15-11}{15-11}\right]\Rightarrow\ Q_{\frac{1}{4}}=86+5\ \Rightarrow\ Q_{\frac{1}{4}}=91$$

4. Cálculo varianza y desviación estándar

a. Varianza

$$\text{Var} = \sigma^2 = \frac{\sum n_i \left(X_i - M\right)^2}{\sum n_i} = \frac{6969{,}6}{60} = 116{,}16$$

b. Desviación Típica

$$\sigma = \sqrt{Var} = \sqrt{116{,}116} = 10{,}77$$

Ejemplo 2

Un estudio sobre los valores de la presión arterial sistólica de estudiantes del área de la salud, arrojó los siguientes datos:

80	90	95	100	110
90	90	100	100	110
90	90	100	105	120
90	90	100	110	120
90	90	100	110	
90	90	100	110	

Con base en esta información es necesario presentar un informe detallado de dicha investigación, para lo cual se necesita calcular lo siguiente:

a) La tabla estadística
b) Los histogramas o gráficas de intervalos vs. fi y Fi.
c) Calcular e interpretar promedio, mediana y moda.

Solución:

Como no dieron el número de intervalos, entonces los calculamos con la siguiente formula

$$K = 1 + 3.3 \, Log \, n$$
$$K = 1 + 3.3 \, Log \, 27$$
$$K = 5.719$$

El rango es

$$R = 120 - 80$$
$$R = 40$$

La amplitud de cada intervalo es

$$I = \frac{R}{K} = \frac{40}{5.719} = 6.99 \cong 7$$

Tabla 22.
Distribución de la presión arterial sistólica

Presión Sistólica	Xi mc	fi	Fi	fiXi	hi	Hi	HI%
80-87	83.1	1	1	83.5	0.037	0.037	37%
87-94	90.5	10	11	905	0.370	0.407	40.7%
94-101	97.5	8	19	780	0.296	0.703	70.7%
101-108	104.5	1	20	104.5	0.037	0.74	74%
108-115	111.5	5	25	557.5	0.185	0.925	92.5%
115-122	118.5	2	27	237	0.074	0.99	99.9%
		$\sum = 2$		$\sum = 2667.5$			

Fuente: Elaboración propia.

La media aritmética es:

$$\overline{X} = \frac{fiXi}{n}$$

$$\overline{X} = \frac{2667.5}{27} = 98.79 \text{ mmHg de presión arterial}$$

$$\overline{X} = 98.79 \text{ mmHg presión sistólica}$$

La moda está dada por:

$$Mo = Li + \mathrm{I}\left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)$$
$$Mo = 87 + 7(0.45)$$
$$Mo = 87 + 3.15$$
$$Mo = 90.15 \ \ mmHg \ \ de\ presión \ \ arterial$$

La mediana está dada por

$$Me = Li + I\left(\frac{\mathrm{N/2 - F_{i-1}}}{\mathrm{F_i - F_{i-1}}}\right)$$

$$Me = 94 + 7\left[\frac{13.5 - 11}{8}\right]$$
$$Me = 94 + 7\ \ (0.312)$$
$$Me = 94 + 2.1875$$
$$Me = 96.18 \ \ mmHg \ \ de\ presión \ \ arterial$$

El 50% de la población presenta una presión arterial sistólica menor o igual al 96.18 mmHg, el otro 50% de la población presenta una presión arterial sistólica mayor o igual al 96.18 mmHg.

Los cuarteles están dados por:

a. Primer cuartil

$$Q_1 = 1\left(\frac{27}{4}\right) = 6.75$$
$$Q_1 = 87 + 7(0.575)$$
$$Q_1 = 87 + 4.025$$
$$Q_1 = 91.025 mmHg \text{ presión arterial Sistólica}$$

Interpretación: El 25% de los estudiantes tienen una presión arterial menor o igual a 91.025 mmHg. En los otros 75% la presión es igual o mayor a 91.025 mmHg.

b. Segundo cuartil

$$Q_2 = 2\left(\frac{27}{4}\right) = 13.5$$
$$Q_2 = 94 + 7\left(\frac{13.5\text{-}11}{8}\right) = 94 + 7\ (0.312)$$
$$Q_2 = 96.18\ mmHg \text{ presión arterial Sistólica}$$

Interpretación: El 50% de los estudiantes tienen una presión arterial igual o menor a 96.18 mmHg. El otro 50% de los estudiantes presentan una presión arterial igual o mayor a 96.18 mmHg.

c. Tercer Cuartil

$$Q_3 = 3\left(\frac{27}{4}\right)20.25$$

$$Q_3 = 108 + 7\left(\frac{20.25\text{-}20}{5}\right) = 108 + 7(0.05)$$

$$Q_3 = 108 + 0.35$$

$$Q_3 = 108.35 \; mmHg \;\; presión \; arterial \; Sistólica$$

Interpretación: El 75% de la población presenta una presión arterial sistólica mayor o igual a 108,35 mmHg. El otro 25% de la población presenta una presión arterial sistólica menor o igual a 108.35 mmHg.

Los gráficos son:

a) Histograma

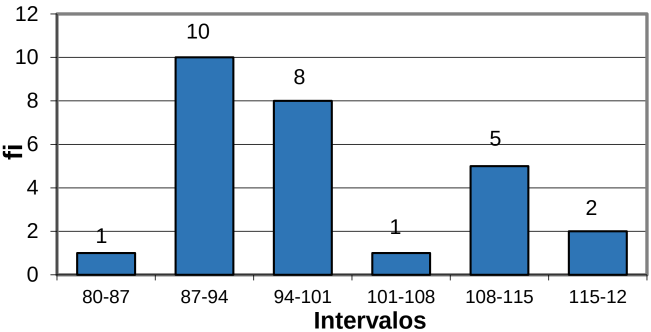

Figura 11. ***Histograma de frecuencia de la presión arterial***

Fuente: Elaboración propia.

b) EL Polígono

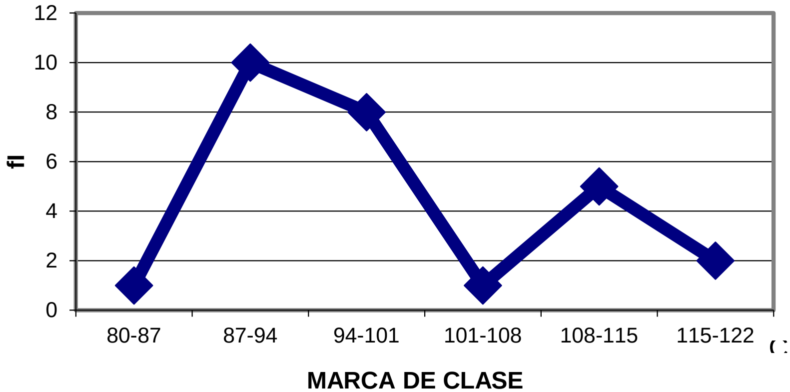

Figura 12. ***Polígono de frecuencia de la presión arterial***

Fuente: Elaboración propia.

Ejemplo 3

Una planta de medicamentos XZ ha recopilado los siguientes datos sobre los volúmenes de ventas del medicamento XZ durante un período de 50 días (Los valores están expresado en 100 libras).

7	11	8	7	10	10	10	6	5	10
10	12	9	9	11	11	7	5	6	9
8	6	10	10	15	15	7	8	8	7
5	5	5	10	8	8	12	8	13	7
8	6	8	14	9	5	5	9	14	8

Construir utilizando 7 intervalos de clase a:

a. La tabla estadística
c) Los histogramas o gráficas de intervalos vs. fi y Fi.
d) Calcular y dar interpretación a: el promedio, la mediana y la moda.

Solución:

Tabla 23.
Distribución en libras del medicamento X

Intervalos	Marca de Clase X_i	Frec. de datos f_i	Frec. Acum.. F_i	Frec. Porc. h_i%	Frec. acum.. H_i%	$X_i.f_i$	$(X_i - \bar{X})^2$	$f_i\,(X_i - \bar{X})^2$
5 - 6.42	5.71	11	11	22%	22%	62.81	9.597	105.567
6.42_1 -7.84	7.13	6	17	12%	34%	42.78	2.815	16.89
784_1 - 9.26	8.55	15	32	30%	64%	128.25	0.066	0.99
9.26_1 –10.68	9.97	8	40	16%	80%	79.76	1.350	10.8
10.68_1-12.1	11.39	5	45	10%	90%	56.95	6.666	33.33
12.1_1-13.52	12.81	1	46	2%	92%	12.81	16.016	16.016
13.52_1 - 15	14.26	4	50	8%	100%	57.04	29.724	118.896
Totales		**40**		**100%**		**440.4**	**66.234**	**302.489**

Fuente: Elaboración propia.

$$a = \frac{Ls - Li}{\# Int} \Rightarrow a = \frac{15-5}{7} \Rightarrow a = \frac{10}{7} \Rightarrow \quad a = 1.42$$

$$\overline{X} = \frac{\sum Xi.fi}{N} \Rightarrow \overline{X} = \frac{440.4}{50} \Rightarrow \overline{X} = 8.808$$

El promedio de libras del medicamento *X* de la planta procesadora es de 8.808 libras

$$Me = Li + a\left[\frac{\frac{N}{2} - F_{i-1}}{F_i - F_{i-1}}\right] \Rightarrow \frac{N}{2} = \frac{50}{2} \Rightarrow \frac{N}{2} = 25 \Rightarrow F_i \geq 25$$

$$Me = 7.84 + 1.42\left[\frac{25-17}{32-17}\right] \Rightarrow Me = 7.84 + 1.42\frac{8}{15} \Rightarrow Me = 7.84 + 0.75 \Rightarrow 8.59$$

La Mediana representa el valor del dato que distribuye uniformemente los datos en forma ordenada, es decir, que la mediana es igual a 8.59.

Moda = 8

La Moda representa los valores que más se repiten dentro de una serie de datos.

Varianza (σ^2) = $\dfrac{\sum\left(Xi - \overline{x}\right)^2 fi}{N} \Rightarrow \sigma^2 = \dfrac{302.489}{50} \Rightarrow \sigma^2 = 6.04978$

Desviación típica $\sigma = \sqrt{6.04978} \Rightarrow \sigma = 2.459$

Cuartiles

$$Q_{\frac{1}{4}} = Li + a\left[\frac{\frac{N}{4} - F_{i-1}}{F_i - F_{i-1}}\right] \Rightarrow \frac{N}{4} = \frac{50}{4} \Rightarrow \frac{N}{4} = 12.5 \Rightarrow F_i \geq 12.5$$

$$Q_{\frac{1}{4}} = 6.42 + 1.42\left[\frac{12.5-11}{17-11}\right] \Rightarrow Q_{\frac{1}{4}} = 6.42 + 1.42\frac{1.5}{6} \Rightarrow Q_{\frac{1}{4}} = 6.42 + 0.355$$

$$\Rightarrow Q_{\frac{1}{4}} = 6.775$$

$$Q_{\frac{2}{4}} = Li + a\left[\frac{\frac{2N}{4} - F_{i-1}}{F_i - F_{i-1}}\right] \Rightarrow \frac{2N}{4} = \frac{2(50)}{4} = 25 \Rightarrow \mathrm{F_i} \geq 25 \Rightarrow$$

$$Q_{\frac{2}{4}} = Li + a\left[\frac{\frac{2N}{4} - F_{i-1}}{F_i - F_{i-1}}\right] = Me = Li + a\left[\frac{\frac{N}{2} - F_{i-1}}{F_i - F_{i-1}}\right]$$

$$Q_{\frac{3}{4}} = Li + a\left[\frac{\frac{3N}{4} - F_{i-1}}{F_i - F_{i-1}}\right] \Rightarrow \frac{3N}{4} = \frac{3(50)}{4} \Rightarrow \frac{3N}{4} = \frac{150}{4} = 37.5 \Rightarrow \mathrm{F_i} \geq 37.5$$

$$Q_{\frac{3}{4}} = 9.26 + 1.42\left[\frac{37.5 - 32}{40 - 32}\right] \Rightarrow Q_{\frac{3}{4}} = 9.26 + 1.42\frac{5.5}{8} \Rightarrow Q_{\frac{3}{4}} = 9.26 + 0.97$$

$$\Rightarrow Q_{\frac{3}{4}} = 10.23$$

$$Q_{\frac{4}{4}} = Li + a\left[\frac{\frac{4N}{4} - F_{i-1}}{F_i - F_{i-1}}\right] \Rightarrow \frac{4N}{4} \Rightarrow \mathrm{N} = 50 \quad Q_{\frac{4}{4}} = 13.52 + 1.42\left[\frac{50 - 46}{50 - 46}\right] \Rightarrow$$

$$Q_{\frac{4}{4}} = 13.52 + 1.42\frac{4}{4} \Rightarrow Q_{\frac{4}{4}} = 13.52 + 1.42 \Rightarrow Q_{\frac{4}{4}} = 14.94$$

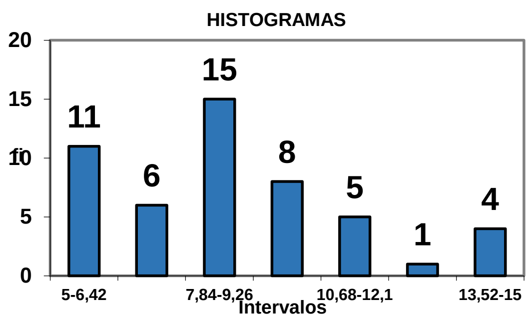

Figura 13. *Histograma de frecuencia de las libras del medicamento X*

Fuente: Elaboración propia.

El intervalo de confianza es: ($\overline{X} - \sigma$; $\overline{X} + \sigma$) = (6.349; 11.267)

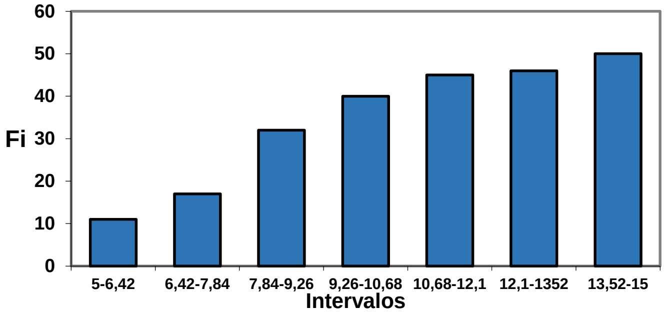

Figura 14. ***Histograma de frecuencia acumulada de las libras del medicamento X***
Fuente: Elaboración propia.

Ejemplo 4

Un estudio sobre el estado de la hemoglobina a un grupo de pacientes arrojó los siguientes datos medidos en gramo por ciento (gm%).

14	11	18	17	10	13	10	15	15	13
15	12	19	19	11	11	17	18	16	19
17	16	16	16	15	15	17	18	18	17
15	15	15	15	18	18	12	18	13	17
18	16	18	14	19	15	15	19	14	18

Los valores reales buenos (V.R.B),son:

Hombres: 14 – 18 gm%
Mujeres: 12 – 16 gm%

Construir utilizando 7 intervalos de clase para:

a. La tabla estadística; si son todos hombres, cuántos y qué porcentaje está en la franja normal, pero si la muestra es solo de mujeres, cuántas de ellas y qué porcentaje están en la franja normal.
b. Los histogramas o gráficas de intervalos vs. fi y Fi.
c. Calcular y dar interpretación para el promedio, la mediana y la moda.

Solución:

Tabla 24.
Distribución del estado de la hemoglobina

Intervalos	Marca de Clase X_i	Frec. de datos f_i	Frec. Acum.. F_i	Frec. Porc. $h_i\%$	Frec. acum.. $H_i\%$	$X_i.f_i$	$(X_i - \bar{x})^2$	$f_i (X_i - \bar{x})^2$
10 - 11.28	10.64	5	5	10%	10%	53.2	23.7169	118.5845
11.28_1 – 12.56	11.92	2	7	4%	14%	23.84	128881	25.7762
12.56_1 – 13.84	13.2	3	10	6%	20%	39.6	5.3361	16.0083
13.84_1 – 15.12	14.48	14	24	28%	48%	202.72	1.0609	14.8526
15.12_1-16.4	15.76	5	29	10%	58%	78.8	0.0625	0.3125
16.4_1-17.68	17.04	6	35	12%	70%	102.24	2.3409	14.0454
17.68_1 - 19	18.34	15	50	30%	100%	275.1	8.0089	120.1335
Totales		**50**		**100%**		**775.5**	**53.4143**	**309.713**

Fuente: Elaboración propia.

Los Hombres: 14 + 5 + 6 + 15 = 40
Las Mujeres: 3 + 14 + 5 = 22

Los que se encuentran en la franja normal son los hombres y su porcentaje es de:

28% + 10% + 12% + 30% = 80

$$a = \frac{Ls - Li}{\# Int} \Rightarrow a = \frac{19 - 10}{7} \Rightarrow a = \frac{9}{7} \Rightarrow \quad a = 1.28$$

$$\overline{X} = \frac{\sum Xi.fi}{N} \Rightarrow \overline{X} = \frac{775.5}{50} \Rightarrow \overline{X} = 15.51$$

El promedio del estado de la hemoglobina del grupo de pacientes es de 15.51 gm%

$$Me = Li + a\left[\frac{\frac{N}{2} - F_{i-1}}{F_i - F_{i-1}}\right] \Rightarrow \frac{N}{2} = \frac{50}{2} = 25 \Rightarrow F_i \geq 25$$

$$Me = 15.12 + 1.28\left[\frac{25 - 24}{29 - 24}\right] \Rightarrow Me = 15.12 + 1.28\frac{1}{5} \Rightarrow$$

$$Me = 15.12 + \frac{1.28}{5} \Rightarrow Me = 15.12 + 0.256 \Rightarrow Me = 15.376$$

La Mediana representa el valor del dato que distribuye uniformemente los datos en forma ordenada, es decir, que la mediana es igual 15.376

Moda = 18

Representa los valores que más se repiten dentro de una serie de datos.

Varianza (σ^2) = $\dfrac{\sum\left(Xi-\overline{x}\right)^2 fi}{N} \Rightarrow \sigma^2 = \dfrac{309.713}{50} \Rightarrow \sigma^2 = 6.19426$

Desviación típica $\sigma = \sqrt{6.19426} \Rightarrow \sigma = 2.488$

Cuartiles

$$Q_{\frac{1}{4}} = Li + a\left[\frac{\frac{N}{4} - F_{i-1}}{F_i - F_{i-1}}\right] \Rightarrow \frac{N}{4} = \frac{50}{4} = 12.5 \Rightarrow F_i \geq 12.5$$

$$Q_{\frac{1}{4}} = 13.84 + 1.28\left[\frac{12.5-10}{24-10}\right] \Rightarrow Q_{\frac{1}{4}} = 13.84 + 1.28\frac{2.5}{14} \Rightarrow$$

$$Q_{\frac{1}{4}} = 13.84 + 0.228 \Rightarrow Q_{\frac{1}{4}} = 14.068$$

$$Q_{\frac{2}{4}} = Li + a\left[\frac{\frac{2N}{4} - F_{i-1}}{F_i - F_{i-1}}\right] \Rightarrow \frac{2N}{4} = \frac{2(50)}{4} = 12.5 \Rightarrow F_i \geq 12.5 \Rightarrow$$

$$Q_{\frac{2}{4}} = Li + a\left[\frac{\frac{2N}{4} - F_{i-1}}{F_i - F_{i-1}}\right] = Me = Li + a\left[\frac{\frac{N}{2} - F_{i-1}}{F_i - F_{i-1}}\right]$$

$$Q_{\frac{3}{4}} = Li + a\left[\frac{\frac{3N}{4} - F_{i-1}}{F_i - F_{i-1}}\right] \Rightarrow \frac{3N}{4} = \frac{3(50)}{4} = \frac{150}{4} =$$

$$37.5 \Rightarrow F_i \geq 37.5$$

$$Q_{\frac{3}{4}} = 17.68 + 1.28\left[\frac{37.5-35}{50-35}\right] \Rightarrow Q_{\frac{3}{4}} = 17.68 + 1.28\frac{2.5}{15} \Rightarrow$$

$$Q_{\frac{3}{4}} = 17.68 + 0.213 \Rightarrow Q_{\frac{3}{4}} = 17.893$$

$$Q_{\frac{4}{4}} = Li + a\left[\frac{\frac{4N}{4} - Fi - 1}{Fi - Fi - 1}\right] \Rightarrow \frac{4N}{4} = \frac{4(50)}{4} =$$

$$50 \Rightarrow Q_{\frac{4}{4}} = 17.68 + 1.28\left[\frac{50-35}{50-35}\right] \Rightarrow$$

$$Q_{\frac{4}{4}} = 17.68 + 1.28\frac{15}{15} \Rightarrow Q_{\frac{4}{4}} = 17.68 + 1.28 \Rightarrow Q_{\frac{4}{4}} = 18.96$$

Los gráficos de los histogramas de frecuencia relativa y acumulada son:

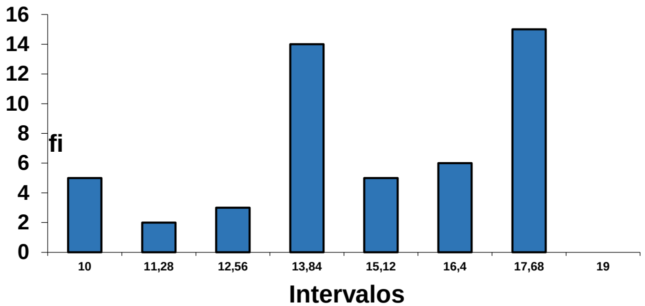

Figura 15. Histograma de frecuencia del estado de la Hemoglobina
Fuente: Elaboración propia.

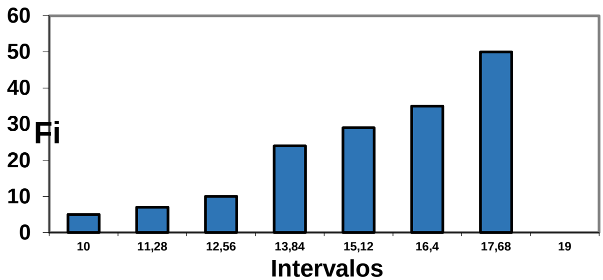

Figura 16. ***Histograma de frecuencia acumulada del estado de la Hemoglobina***
Fuente: Elaboración propia.

AUTOEVALUACIÓN

1. En la siguiente tabla se encuentran registrados los pesos de 40 estudiantes de la Universidad (medidos en libras). Construir una distribución de frecuencia (utilizando 6 clases o intervalos)

138	164	150	144	125	149	157
146	158	140	136	148	152	144
168	126	138	133	119	154	165
146	173	142	135	153	140	135
161	145	135	150	156	145	128

Construir:

a. Un histograma,
b. Un polígono de frecuencia para la distribución de los pesos del problema anterior.
c. Calcular el promedio, la mediana y la moda.

2. Los datos a continuación representan las edades de los pacientes que ingresan al Hospital Rural de Palestina el día 30 de febrero de 1997.

180	175	160	143	142	188	185	156	154	167
187	185	165	159	175	187	184	152	144	141

   a. Construir una distribución de frecuencia con las clases 40-49, 50-59, etc.
   b. Obtener la media muestral con la distribución de frecuencia.
   c. Calcular la media muestral con los datos brutos.
   d. Comparar (b) y (c) y comente su respuesta.

3. Un Almacén anuncia: “Sí nuestros precios promedio no son iguales ó más bajos que los de cualquier otro almacén, le regalamos la mercancía”. Un cliente llega un día al almacén y pone sobre el mostrador facturas de 6 artículos que compró a un competidor pagando un precio promedio menor. He aquí el valor de las facturas: \$1.29, \$2.97, \$3.49,\$5.00, \$7.50, \$10.95.

Los precios que esas mismas 6 mercancías tenían en el almacén son: \$1.35, \$2.89, \$3.19, \$4.98, \$7.59 y \$11.50. El dueño dijo al cliente: “Mi anuncio se refiere al promedio ponderado del precio de esos artículos. Nuestro promedio es menor porque nuestras ventas de ellos han sido:

7	9	12	8	6	3

¿Está el dueño del almacén en problemas o saliendo de ellos al hablar de promedios ponderados? Justifique.

4. Una empresa textil ha mostrado los siguientes incrementos porcentuales en el capital durante los últimos 5 años.

1982	1983	1984	1985	1986
5%	10.5%	9.0%	6.0%	7.5%

¿Cuál es el incremento porcentual promedio en el capital durante el período de cinco años?

5. El número de horas en que se ve televisión por familia, y el horario con mayor número de telespectadores, son dos factores que influyen en los precios de la publicidad por televisión. Una muestra aleatoria de 25 familias en Valledupar, produjo las siguientes estimaciones del tiempo que se dedica a ver televisión por familia:

3.0	6.0	7.5	15.0	12.0
6.5	8.0	4.0	5.5	6.0
5.0	12.0	1.0	3.5	3.0
7.5	5.0	10.0	8.0	3.5
9.0	2.0	6.5	1.0	5.0

a. Calcular la media muestral $\bar{x}$ y la desviación estándar muestral ***s*.**
b. Encontrar el porcentaje de horas de ver televisión por familia que caen en el intervalo $\bar{x} \pm \mathbf{2}s$. Comparar su respuesta con el porcentaje correspondiente dado por la regla empírica.

6. A continuación se relacionan los promedios de calificaciones de 25 estudiantes de la Universidad de Santander del segundo semestre en Ingeniería de Sistemas.

2.6	1.8	2.6	3.7	1.9
2.1	2.7	3.0	2.4	2.3
3.1	2.6	2.6	2.5	2.7
2.7	2.9	3.4	1.9	2.3
3.3	2.2	3.5	3.0	2.5

a. Trazar un histograma de frecuencias relativas.
b. Calcular $\bar{x}$ y s.
c. Determinar los intervalos $(\bar{x} \pm s)$, $(\bar{x} \pm \mathbf{2}s)$; y $(\bar{x} \pm \mathbf{3}s)$
d. Hallar las cantidades de observación que caen en los intervalos. ¿Concuerdan estos números con el teorema de Chebyshev? ¿Y con la regla empírica?.

7. La duración promedio de los anuncios o avisos comerciales televisivos en el canal Telecaribe es de 75 segundos con una desviación estándar de unos 20 segundos. Al suponer que los tiempos de duración tienen una distribución aproximadamente normal.

a. ¿Qué cantidad aproximada de comerciales durará menos de 35 segundos?
b. ¿Qué cantidad aproximada de comerciales durará más de 55 segundos?

8. A continuación, se muestra una tabla de distribución de frecuencia de los resultados de un test practicado a 120 obreros que laboran en una empresa, el objetivo del test era medir el nivel de conocimiento que tienen los trabajadores sobre los reglamentos y normas que tiene la empresa. Con base en esta información se solicita:

    a) Calcular la mediana
    b) Encontrar la moda.
    c) Hallar el 70° percentil,
    d) y el tercer cuartil.

Clase	**Intervalo**	f_i
1	76 - 80	4
2	81 - 85	7
3	86 - 90	10
4	91 - 95	13
5	96 - 100	25
6	101 - 105	29
7	106 - 110	12
8	111 - 115	11
9	116 - 120	9

9. Una cadena de supermercados compara los precios por mercancías idénticas en todas sus tiendas de comestibles. A continuación se transcriben los precios a que una libra de pollo de vendió en cada tienda la semana anterior:

\$1.08	\$0.98	\$1.09	\$1.24	\$1.33
\$1.14	\$1.55	\$1.08	\$1.22	\$1.05

a. Calcular el precio de la mediana por libra
b. Determinar el precio medio por libra.
c. ¿Qué valor es la mejor medida de tendencia central de estos datos?

10. Para la distribución de frecuencia que se indica a continuación, determinar:

a. Cuál es la mediana de clase.
b. Cuál número de elemento representa la mediana del elemento.
c. La amplitud de los pasos iguales en la mediana de clase.
d. El valor estimado de la mediana de estos datos.

Clase	Frecuencia
100- 149.5	12
300- 349.5	72
150- 199.5	14
350- 399.5	63
200- 249.5	27
400- 449.5	36
250- 299.5	58
450- 499.5	18

11. A continuación, se relacionan las edades de los automóviles con que trabajó Autos-Chevrolet la última semana:

5	6	3	6	11	7	9
10	2	4	10	6	21	5

a. Determinar la moda de este conjunto de datos.
b. Calcular la media del conjunto de datos.
c. Comparar a) y b) y diga cuál es mejor medida de tendencia central de los datos.

12. La agencia publicitaria PubliValle está evaluando los diferentes medios de promoción para uno de sus clientes, un fabricante de productos de aseo destinados a los ancianos. Uno de los eventos que se piensa realizar es una carrera anual de automóviles. A continuación se da la distribución de edades del público que asistirá a las carreras:

Edad	Frecuencia
15 - 20	5
21 - 26	20
27 - 32	24
33 - 38	30
39 - 44	46
45 - 50	88
51 - 56	103
57 - 62	44
63 - 68	30

13. Los siguientes datos corresponden a los tiempos (en semanas) que han requerido 30 personas para conseguir un nuevo empleo Construir una tabla de frecuencia y calcule:

12,	15,	5,	12,	18,	3,	0,	17,	1,	8
9,	8,	2,	6,	11,	7,	13,	4,	0,	6,
10,	0,	11,	2,	13,	1,	7,	5,	1,	3.

a. La desviación estándar.
b. El rango semi-intercuartilico.
c. ¿Cumple la relación entre la desviación estándar y el rango semi-intercuartilico?
d. ¿A qué se debe la poca aproximación?.

14. Una investigación realizada a 20.000 trabajadores que laboran en el sector oficial revela que el monto promedio de sus salarios es de $600.000 y desviación estándar es de $40.000. Halle la proporción de empleados que tienen salarios:

a. Por debajo de $560.000.
b. Por encima de $600.000.
c. Entre $780.000 y $620.000.

15. Durante 8 años, los precios del producto XY tuvieron un promedio de $80 y una desviación de $12. En el tiempo anterior a los 8 años, el promedio era de $50 con una varianza de $36. ¿En qué tiempo se mantuvo más estable?

16. La demanda diaria del producto ZX en un centro comercial durante un mes fue:

48,	45,	86,	68,	58,	69,	77,	73,	43,	69,
33,	61,	38,	35,	46,	46,	71,	67,	59,	88,
58,	42,	82,	72,	57,	76,	68,	54,	54,	66.

a. Tome $k = 1..5$ y calcule $\overline{\mathrm{x}}$ - ks, $\overline{\mathrm{x}}$ + ks
b. Utilizando la regla de Shebyshev predecir el porcentaje de datos entre $\overline{\mathrm{x}}$ - ks, y $\overline{\mathrm{x}}$ + ks

c. Determinar el valor *k*, para que en el intervalo de extremos $\bar{x} - ks$, y $\bar{x} + ks$ quede al menos el 80% de los datos.

17. El número de cheques cobrados diariamente en 5 sucursales de un banco durante el mes anterior tuvo la siguiente distribución de frecuencia:

Clase	Frecuencia
0 - 299	20
300 - 499	23
500 - 699	27
700 - 899	52
900 - 999	28

El gerente del banco BB, conoce que una desviación estándar en el cobro de los cheques de más de 200 cheques diarios crea problemas de organización y dotación de personal en cada una de las sucursales debido a un exceso de trabajo no uniforme. ¿Debe el gerente del banco BB preocuparse por estos movimientos?

18. El administrador de un Hospital de Valledupar realizó una encuesta sobre el tiempo (en días) que 300 pacientes escogidos al azar permanecen después de ser sometidos a una operación. He aquí los datos:

Permanencia en el Hospital (en días)	1 – 3	4 – 6	7 – 9	10 – 12	13 – 15	16 – 18	19 – 21	22 – 24
Frecuencia	25	84	53	23	12	6	3	2

Calcular la media y la desviación estándar.

18. Con los siguientes datos de la población, calcular la desviación absoluta promedio, la variancia y la desviación estándar. ¿Qué indican las respuestas acerca del comportamiento de los costos de combustible de la calefacción? ¿Cuál es el costo promedio de combustible de calefacción por galón en ocho departamentos?

\$1.65	\$1.73	\$1.66	\$1.81
\$1.79	\$1.68	\$1.74	\$1.70

19. Un ingeniero realizó una simulación sobre la forma en que unos usuarios llenan un Test de coeficiente intelectual. Para correr y probar la simulación, digitó en la computadora 18 formas diferentes de un test del coeficiente intelectual.

Valor del CI

134	136	137	138	138
143	144	144	145	146
146	146	147	148	153

Calcular la media y la desviación estándar de las puntuaciones del coeficiente intelectual.

20. La empresa de Consultores Valle y Asociados cuenta con los siguientes registros que indican el número de días en que sus 8 consultores se presentaron a trabajar:

212	220	230	210	228
229	231	219	221	222

a. Sin calcular el valor de ninguna de las medidas anteriores, ¿cuál cree usted que aportará más información sobre esta distribución: el intervalo, la desviación absoluta promedio, la desviación estándar?
b. Considerando la dificultad y el tiempo en el cálculo de cada una de las medidas que repasó en la parte a), ¿cuál es la mejor a su juicio?
c. ¿Qué hará que usted cambie su opinión respecto a su elección?

21. Calcular la media de la muestra, la mediana, la moda, el rango, la varianza y la desviación típica:

a. Unos estudiantes del primer semestre del programa de ingeniería industrial de la U faltaron a la institución, el número de días que fallaron durante el semestre académico son:

2	9	3	3	4
10	3	4	6	3
5	8	10	9	7

b. Las distancias (medidas en cuadras) a que viven 10 empleados de la UA respecto su lugar de trabajo son:

11	20	12	11	4
8	16	5	7	13

22. Si se tienen los datos 2 4 6 8 10 12

    Compruebe que $(x - x) = 0$

23. Sean las dos sucesiones X y Y apareadas así:

X:	2	4	6	8	10
Y:	1	2	3	4	5

a. Hallar la media de W, al ser $W=X - Y$
b. Determinar la media de W, al ser $W=2X + 2Y$

CAPITULO IV

TÉCNICAS DE CONTAR

En este capítulo se desarrollaran algunos métodos para determinar sin enumeración directa el número de resultados posibles de un experimento particular o el número de elementos de un conjunto particular. Tales técnicas son conocidas algunas veces con el nombre de *análisis combinatorio*.

1. ANÁLISIS COMBINATORIO

Cuando el número de puntos o datos muéstrales no es grande, la cuenta directa de los puntos del muestreo necesarios para calcular las probabilidades no es complicada. Sin embargo, cuando la cuenta directa tiene dificultades hace que los procesos se vuelvan tediosos o imposibles de calcular. En tales casos se emplea el *análisis combinatorio,* que podría decirse que es una *forma sofisticada* de contar.

Principio fundamental del Conteo

Si un suceso o evento se puede realizar de n_1 maneras distintas, y si se continua con el procedimiento ejecutado, en un segundo suceso o evento se podría realizar de n_2

formas diferentes, y si, después de haberse generados, un tercer suceso o evento se puede realizar de n_3 maneras distintas, y así continuamente; entonces se puede afirmar que el número de maneras en que los suceso o eventos logran realizarse en el orden indicado está dado por el producto $n_1.n_2.n_3\ldots$

Notación Factorial

El producto sucesivo de números enteros desde 1 hasta n inclusive, es utilizado con mucha frecuencia en los cálculos matemáticos y son denotados por el símbolo $n!$ (que es ¨n factorial¨):

$$n! = 1.2.3\ldots(n-2).(n-1).n$$

Es importante definir que $0! = 1$

2. PERMUTACIONES

Una ordenación de un conjunto de n objetos en un orden dado se llama *permutación* de objetos (tomados todos a la vez). Una ordenación de un número de r de dichos objetos, $r \leq n$, en un orden dado se llama *permutación* r o *permutación de n objetos tomados r a la vez.*

El número de permutaciones de n objetos tomados r a la vez la denotamos por

$$P(n, r) = n.(n-1).(n-2)\ldots(n-r+1) = \frac{n!}{(n\text{-}r)!}$$

Permutaciones con repeticiones

Con frecuencia se necesita el número de permutaciones de objetos, de los cuales algunos son iguales, como se indica a continuación, usando la fórmula general.

Teorema:

El número de permutaciones de n objetos de los cuales n_1 son iguales, n_2 son iguales..., n_r son iguales, es:

$$\frac{n\,!}{n_1!\,.\,n_2!\,\ldots\,n_r!}$$

Ejemplo:

Si un hombre tiene 2 camisas y 4 corbatas entonces tiene 2 * 4 = 8 maneras de escoger una camisa y luego una corbata.

El *diagrama árbol* que se muestra en la siguiente figura, es empleado con mucha frecuencia en la solución de problemas basado en el principio anterior.

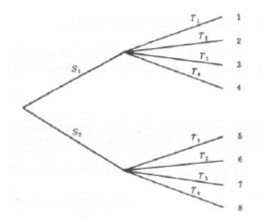

Figura 17. *Diagrama de árbol*
Fuente: Elaboración propia.

Solución:

Si decimos que los suéteres se representan por S1 y S2. Y los pantalones por *T1, T2, T3 y T4; l*as diferentes maneras de escoger un suéter y luego un pantalón se muestran en el diagrama árbol.

Ejemplo:

La placa de un automóvil es diseñada con dos letras distintas seguidas de tres dígitos o números con la condición de que el primero no es cero. ¿Determine el número de placas diferentes que pueden fabricarse?

Solución:

Se puede observar que la primera letra tiene 26 maneras diferentes de ser elegida (suponiendo el alfabeto ingles de 26 letras), la segunda letra tendrá 25 opciones de ser elegida (puesto que la letra grabada de primera no puede seleccionarse otra vez), de igual forma para el primer número o dígito hay nueve opciones entre los números, o sea nueve maneras y para cada uno de los otros dos dígitos 10 maneras. Por tanto, puede grabarse:

$26*25*9*10*10=585000$ Placas diferentes con la condición dada.

Ejemplo:

Encuentre el número de permutaciones de 6 elementos o letras, como son *a, b, c, d, e, f* tomando solo tres a la vez. Otra forma de comprender este problema es calcular el número de palabras de tres letras distintas que pueden crearse con el conjunto de letras dado.

Solución:

Con fórmula:

$$P(6,\ 3) = {}_6P_3 = \frac{6!}{(6-3)!} = 120$$

Con técnica es:

$$6 * 5 * 4 = 120$$

Ejemplo:

¿Calcular el número de señales diferentes, con 8 banderas ubicadas en línea, que pueden armarse usando 4 banderas de color rojo, 3 banderas blancas y una bandera de color azul?

Solución:

Calculemos primero el número de permutaciones de 8 objetos con la condición de que 4 que son iguales (las banderas de color rojo) y 3 también son iguales (las banderas blancas). Usando el teorema de permutación con repetición tenemos:

$$\frac{8!}{4!*3!} = 280 \qquad \text{Señales diferentes}$$

Ejemplo:

El número de permutaciones diferentes de las 11 letras de la palabra *MISSISSIPPI,* que consiste de 1M, 4I, 4S y 2P es:

$$\frac{11!}{1!\ 4!\ 4!\ 2!} = 34650$$

Ejemplo:

¿De cuántas maneras diferentes pueden ordenarse 5 bolas en una fila?

Solución:

Debemos ordenar 5 bolas usando una fila de 5 posiciones, de la siguiente manera: La primera posición se podrá ocupar usando una de las 5 bolas, es decir, que existen 5 opciones de llenar la primera casilla o posición. Continuando con la posición 2 esta tendrá 4 opciones de ser llenada. Luego quedan solo 3 maneras de llenar la tercera casilla, 2 opciones de llenar la cuarta posición, y finalmente sólo 1 manera de llenar la última casilla. Esto es:

El número de ordenaciones de las 5 bolas en una fila es = 5, 4. 3. 2. 1 = 5! = 120

Cuando tengamos que ordenar n objetos diferentes en una fila usaremos la fórmula: $n\ (n - l)\ (n - 2)' \ldots 1 = n!$.

Ejemplo:

¿Calcular el número de maneras en que pueden sentarse 10 personas en una silla que sólo tiene 4 puestos disponibles?

Solución:

El primer puesto que se puede ocupar con cualquiera de las 10 personas tiene 10 opciones, el segundo puesto tendrá 9 maneras de ocuparse, habrá 8 opciones para ocupar el tercer puesto y 7 para ocupar el cuarto. Esto es:

EI número de ordenaciones de 10 personas tomadas de 4 en 4 es = 10. 9. 8. 7 = 5040

De aquí que el número de ordenaciones de n objetos diferentes escogiendo solo de r en r está dado por: $n(n\text{-}1)\ldots(n - r + 1)$.

Ejemplo:

Hallar el valor de (a) ${}_8P_3$. (b) ${}_6P_4$, (e) ${}_{15}P_1$, (d) ${}_3P_3$

a. ${}_8P_3 = 8 * 7 * 6 = 336$
b. ${}_6P_4 = 6 * 5 * 4 * 3 = 360$
c. ${}_{15}P_1 = 15$
d. ${}_3P_3 = 3 * 2 * 1 = 6$

Ejemplo:

Se desea que se sienten cinco hombres y cuatro damas en una fila, de tal manera que las mujeres ocupen las sillas pares. ¿Calcule las formas en que pueden sentarse?

Solución:

Se tiene que los cinco hombres se pueden sentar ${}_5P_5$ formas, mientras que las damas pueden sentarse de ${}_4P_4$ formas. La forma de ordenación de los hombres puede asociarse con cada ordenación de las damas. Así pues,

El número de formas en que pueden sentarse es = ${}_5P_5 * {}_4P_4 = 5!\ 4! = (120)(24) = 2880$

Ejemplo:

¿Cuántos números de cuatro (4) cifras pueden formarse con los 10 dígitos 0, 1, 2, 3, ..., 9 si (*a*) repitiendo los números, (*b*) sin repetir los números, (*c*) si el último número es cero y no pueden repetirse los números?

Solución:

a. La primera cifra tiene 9 opciones (puesto que el 0 no tiene valor). La segunda, tercera y cuarta pueden ser cualquiera de las 10. Entonces 9 10 10 10 = 9000 son los números que se pueden armar.
b. El primer dígito puede tener 9 opciones (el 0 no).

El segundo dígito tiene 9 opciones (no puede ser la que ocupó el primer puesto).
El tercer digito cuenta con 8 opciones (no pueden ser ninguna de las que ocupan los dos primeros puestos).
La cuarta cuenta con 7 opciones (no pueden ser ninguna de las que ocupan los tres primeros puestos).
Luego se tiene que 9 * 9 * 8 * 7 = 4536 son las cantidades de números que pueden formarse.

Otro método.

El primer número puede ser cualquiera entre 9 y los tres restantes pueden elegirse de ${}_9P_3$ formas.

Entonces 9 ${}_9P_3$ = 9 * 9 * 8 * 7 = 4536 son los números que pueden formarse.

c. La primera cifra se puede escoger de 9 maneras, la segunda cifra tiene 8 formas de ser escogida y la tercera tiene 7 opciones. Entonces 9 * 8 * 7 = 504 son los números que pueden formarse.

Otro método.

El primer dígito se puede seleccionar de 9 maneras y los dos dígitos siguientes se pueden escoger de ${}_8P_2$ formas. Entonces 9 * ${}_9P_3$ = 9 * 8 * 7 = 504 será el número pedido.

Ejemplo:

Cuatro (4) libros distintos de matemáticas, seis (6) diferentes de física y dos (2) diferentes de química se colocan en un estante. ¿De cuántas formas distintas son posibles ordenados si (a) los libros por tema o asignatura deben estar todos juntos, (b) solamente los libros de matemáticas deben agruparse juntos??

Solución:

a. Como los libros de matemáticas van juntos, solo pueden ordenarse entre ellos de la forma 4P4 = 4! formas, los libros de física tienen ${}_6P_6 = 6!$ formas, mientras que los libros de química se ordenan de forma ${}_2P_2 = 2!$ y los tres grupos por tema se ordenan de ${}_3P_3 = 3!$ formas.

   Luego, el número de ordenaciones de los libros es = 4! 6! 2! 3! = 207 360

b. Hay que tomar los cuatro libros de matemáticas como si fuera un solo libro. Por lo tanto, se tienen. 9 libros que podríamos ordenar de ${}_9P_9 = 9!$ formas. En todos estos casos los libros de matemáticas están juntos. Pero los libros de matemáticas pueden ordenarse entre ellos de ${}_4P_4 = 4!$ formas.

Por ultimo tenemos que el número de ordenaciones es = 9! 4! = 8709120

Ejemplo:

Se desea ordenar en una fila cinco bolas de color rojo, dos bolas de color blanco y tres bolas de color azul. ¿de cuántas formas es posible ordenarlas?

Solución:

Se supone que hay N' diferentes ordenaciones. Multiplicando N por el número de ordenaciones *(a)* de las 5 bolas rojas entre sí, *(b)* de las 2 bolas blancas entre sí y *(c)* de las 3 bolas azules entre sí (es decir, multiplicando N por 5! 2! 3!) se obtiene. El número de ordenaciones de 10 bolas si todas ellas fuesen distintas, es decir, 10!

Entonces $(5!\ 2!\ 3!)N = 10!$ y $N = 10!/(5!\ 2!\ 3!)$,

Ejemplo:

¿Calcule el número de maneras en que pueden sentarse siete personas alrededor de una mesa, si: *(a)* se sientan de cualquier forma, y (b) dos de las personas invitadas no deben estar juntas?

Solución:

a. Piense que una de las personas se puede sentar en cualquier parte alrededor de la mesa. Entonces los seis restantes se podrán sentar de 6! = 720 formas, que es el total de casos que se dan al sentar a siete personas alrededor de una mesa circular.

b. Ahora piense en las dos personas que no pueden estas juntas como una sola. Entonces tendremos solo seis personas para sentarse en círculo, que se puede resolver de 5! formas diferentes. Pero estas dos personas consideradas como una sola si se pueden ordenar de 2! formas. Luego, el número de ordenamiento de 6 personas sentadas alrededor de una mesa con dos de ellas que no pueden ser sentadas juntas es de 5! 2! = 240.

Como en el item *(a),* se tiene como el número de formas en que seis personas pueden sentarse alrededor de una mesa, con la condición de que dos de ellas no pueden estar sentadas juntas es 720-240 = 480 formas.

3. COMBINACIONES

Si se tiene una colección de *n* elementos u objetos. Una combinación de estos *n* elementos u objetos seleccionando solo *r* a la vez, es un subconjunto de *r* elementos. Es decir, que es una combinación de la selección de *r* o de *n* objetos, aquí el orden de los elementos u objetos no se tiene en consideración.

$$C(n,\mathrm{r}) = \frac{n!}{(n-r)!\ r!} \quad \text{o también} \quad C(n,\mathrm{r}) = \frac{P(n,\mathrm{r})}{r!}$$

Ejemplo

Las combinaciones de las letras *a, b, c, d* tomadas 3 la vez son:

$$C(4,3)=\frac{4!}{(4-3)!\ 3!}=4$$

Las cuales son:

$\{a, b, c\}$, $\{a, b, d\}$, $\{a, c, d\}$, $\{b, c, d\}$ o simplemente *abc, abd, acd, bcd.*

Obsérvese que las combinaciones siguientes son iguales: *abc, acb, bac, bca, cab, cba* ó sea, cada una representa el mismo conjunto $\{a, b, c\}$.

Ejemplo:

Hallar el valor de *(a)* ${}_7C_4$, *(b)* ${}_6C_5$ *(c)* ${}_4C_4$

Solución:

a. $${}_7C_4=\frac{7!}{4!\ 3!}=\frac{7*6*5*4}{4!}=\frac{7*6*5}{3*2*1}=35$$

b. $${}_6C_5=\frac{6!}{5!\ 1!}=\frac{6*5*4*3*2}{5!}=6$$

c. $${}_4C_4=\frac{4!}{0!\ 4!}=\frac{4*3*2*1}{1*4*3*2*1}=1$$

Ejemplo:

El número de maneras en las cuales 3 cartas pueden escogerse o seleccionarse de un total de 8 cartas diferentes es:

$$ {}_8C_3 = C(8,3) = \binom{8}{3} = \frac{8*7*6}{3*2*1} = 56 $$

Ejemplo:

¿Calcule las maneras de seleccionar un comité, conformado por tres caballeros y dos damas, de un grupo de siete caballeros y cinco damas?

Solución:

De los 7 caballeros se pueden escoger 3 de la forma $\binom{7}{3}$, y de las 5 damas se pueden escoger 2 de la manera $\binom{5}{2}$

Por lo tanto, el comité puede escogerse de:

$$ \binom{7}{3} * \binom{5}{2} = \frac{7*6*5}{3*2*1} * \frac{5*4}{1*2} = 350 \textit{ maneras} $$

Ejemplo:

¿De cuántas formas puede elegirse una comisión de 5 personas de entre 9 personas?

Solución:

$$ {}_9C_5 = \binom{9}{5} = \frac{9!}{5!\ 4!} = \frac{9*8*7*6*5}{5!} = 126 $$

Ejemplo:

Se tiene un grupo de profesionales conformado por cinco matemáticos y siete físicos, y se decide formar un comité de dos matemáticos y tres físicos. Determine el número de formas como puede conformarse, si *(a)* pueden pertenecer al comité cualquier Matemático y Físico. *(b)* Si se desea que un físico determinado pertenezca al comité, y *(c) q*ue dos matemáticos determinados deben estar dentro del comité.

Solución:

a. De los 2 matemáticos, 5 pueden elegirse con una combinación de ${}_5C_2$. Y de los 3 físicos de 7 pueden elegirse con una combinación de ${}_7C_3$ formas.

Número total de selecciones posibles = ${}_5C_2 * {}_7C_3 =$ 10 * 35 = 350

b. 2 matemáticos de un total de 5 pueden elegirse de ${}_5C_2$ formas. 2 físicos restantes de un total de 6 pueden elegirse de ${}_6C_2$ formas.

Número total de selecciones posibles = ${}_5C_2 * {}_6C_2 =$ = 10 * 15 = 150

c. 2 matemáticos de un total de 3 pueden elegirse de ${}_3C_2$ formas. 3 físicos de un total de 7 dan ${}_7C_3$ formas.

Número total de selecciones posibles = ${}_3C_2 * {}_7C_3 =$ 3 * 35 = 105

Ejemplo:

¿Cuántas ensaladas se pueden preparar con lechuga, escarola, endibia, berro y achicoria?

Solución:

Cada una de las verduras para los platos se puede considerar de dos maneras (se escoge o se rechaza para el plato). Puesto que cada una de las dos maneras de considerar una verdura está asociada con dos formas de considerar cada una de las otras verduras, el número de formas de decidir de las cinco verduras es 2^5. Pero las 2^5 maneras incluyen el caso de no seleccionar ninguna verdura. Por tanto, la solución al problema es:

El número de ensaladas es = $2^5 - 1 = 31$

Otro método.

Puede seleccionarse 1 de las 5 verduras, 2 de las 5 verduras, y así sucesivamente. El número requerido de ensaladas es:

$$_5C_1 +_5 C_2 +_5 C_3 +_5 C_4 +_5 C_5 = 5 + 10 + 10 + 5 + 1 = 31$$

Ejemplo:

Con 7 consonantes y 5 vocales diferentes. ¿Cuántas palabras pueden formarse, que consten de 4 consonantes y 3 vocales? No es necesario que las palabras tengan significado.

Solución:

Las 4 consonantes pueden elegirse de $_7C_4$ formas, las 3 vocales de $_5C_3$ formas y las 7 letras resultantes. (4 consonantes y 3 vocales) pueden ordenarse entre sí de $_7C_7$,= 7! formas. Entonces, el número de palabras es:

$$_7C_4 * _5C_3 * 7! = 35 * 10 * 5040 = 1764000$$

Coeficientes Binomiales

Los números de $C(n, r)$ se les llama frecuentemente los *coeficientes binomiales* puesto que provienen de la *expansión binomial.*

$$(x+y)^n = x^n + \binom{n}{1} x^{n-1} y + \binom{n}{2} x^{n-2} y^2 + \cdots + \binom{n}{n} y^n$$

Tienen muchas propiedades.

Ejemplo:

$$\begin{aligned}(x+y)^4 &= x^4 + \binom{4}{1} x^3 y + \binom{4}{2} x^2 y^2 + \binom{4}{3} x y^3 + \binom{4}{4} y^4 \\ &= x^4 + 4x^3 y + 6x^2 y^2 + 4x y^3 + y^4\end{aligned}$$

Aproximación de Stirling a *N!*

Cuando n es muy grande la evaluación de $n!$ no es práctica. En tales casos se utiliza la fórmula aproximada

$$n! \cong \sqrt{2\pi n}\, n^n e^{-n}$$

Las constantes de la formula anterior son:

$e = 2.71828$

$\pi = 3.1416$

El símbolo $\approx$ significa que la relación del lado izquierdo al lado derecho se aproxima a 1 a medida que *n* tienda al infinito, es decir $n \to \infty$. Por esta razón decimos que el lado derecho es una *expansión asintótica* del lado izquierdo. Para un estudio más detallado de la fórmula de Stirling.

4. INTRODUCCIÓN A LA PROBABILIDAD

La *Probabilidad* se encarga de estudiar experimentos basados en lo aleatorio o que son difíciles para determinar. Por ejemplo, si lanzamos un dado, entonces tenemos la certeza de que caerá a una superficie o suelo, pero no podemos afirmar que saldrá un 5. Pero, si repetimos el experimento de volver a lanzar el dado; y llamamos *s* al número de aciertos, esto es, el número de veces en que un 5 aparece, y sea *n* el número de lanzadas. Entonces podemos deducir empíricamente la relación:

$$f = \frac{s}{n},\ \text{llamada}\ \textit{frecuencia relativa.}$$

Con tendencia a estabilizarse a la larga, aproximándose a un límite. Esta posibilidad es la base de la teoría de la probabilidad (Walpole, Myers y Myers, 1999).

La teoría de la probabilidad define una estructura o modelo matemático que permite asociar eventos como los anteriores con un experimento. El modelo matemático empleado para un experimento solo depende de la aproximación de las posibilidades asignadas con la frecuencia real relativa. Este análisis da origen a los problemas de comprobación y confiabilidad que son elementos fundamentales en la estadística.

La teoría moderna de la probabilidad está basada en axiomas, esto implica que las probabilidades de los eventos pueden ser arbitrarias, sin olvidar que estas deben satisfacer a los axiomas de la probabilidad moderna, mientras que la teoría clásica corresponde y se rige solo por los *espacios equiprobables*.

El Concepto de Probabilidad

En cualquier experimento de tipo aleatorio está presente la incertidumbre sobre si un suceso o evento específico ocurrirá o no. La cual podemos concebir como la medida de la *oportunidad* o de la *probabilidad*. Si estamos seguros que el evento ocurrirá podemos decir que su probabilidad es 100% ó 1, pero si se está seguro de que el evento no ocurrirá entonces su probabilidad es cero. Por ejemplo, si la probabilidad que suceda un evento es de ¼, se puede decir que hay un 25% de oportunidad que ocurra y existe un 75% de que no ocurra.

Existen dos procedimientos importantes por medio de los cuales podemos obtener estimativos para la probabilidad de un suceso.

1. *Enfoque clásico o a priori*. Si un evento o suceso ocurre en h maneras de un número total de n formas posibles, todos con igual factibilidad, entonces la probabilidad del suceso es h/n.

   Ejemplo:

   Si deseamos la probabilidad de sacar una cara en un lanzamiento de una moneda. Como existen dos maneras (cara o sello) como resultado de la moneda. De estas dos maneras una puede aparecer, Entonces la probabilidad de sacar una cara es 1/2.

2. *Enfoque como frecuencia relativa o a posteriori*. Si después de realizar n repeticiones en un experimento, con n muy grande, y un suceso ocurre h veces, entonces la probabilidad del suceso es h/n. la cual es conocida como: la *probabilidad empírica* del suceso. También se llama la *probabilidad empírica* del suceso.

   Ejemplo:

Si lanzamos una moneda 1000 veces y hallamos que 532 veces resultan caras estimamos que la probabilidad de una cara es 532/1000 = 0.532.

Ambos enfoques el clásico y el de frecuencia presentan serias dificultades, el primero debido a la vaguedad de las palabras "igualmente factibles" y el segundo debido a la vaguedad incluida en un "número muy grande". A causa de estas dificultades los matemáticos se han orientado a un *enfoque axiomático* utilizando conjuntos.

Axiomas de Probabilidad

Si S es un espacio muestral, y ε es la colección de eventos, siendo P una función de valores reales bien definida en ε. Entonces P será una función de probabilidad, y $P(A)$ es la probabilidad del evento y A si se cumplen los siguientes axiomas:

1. Para todo evento $A, \quad 0 \leq P(A) \leq 1$
2. $P(\phi) = 0$
3. $P(S) = 1$
4. Además, si A_1, A_2, A_3,… es una serie de eventos mutuamente excluyentes, entonces

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + P(A_3) + \cdots$$

TEOREMAS

1. Si A^c es el complemento del evento A, entonces $P(A^c) = 1 - P(A)$
2. Si $A \subset B$, entonces $P(A) \leq P(B)$
3. Si A y B son dos eventos, entonces $P(A \setminus B) = P(A) - P(A \cap B)$
4. Si A y B son dos eventos, entonces $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Ejemplo

En un hipódromo se llevaron tres caballos pura sangre, A, B y C, los cuales por su trayectoria y carrera se sabe que: el caballo A tiene doble posibilidad de ganar que el caballo B; y B el doble de ganar que el caballo C. Calcule las probabilidades de ganar de cada uno de ellos, es decir: P(A), P(B), P(C)

Solución:

Sea $P(C) = p$; siendo B con el doble de probabilidad de ganar que C, $P(B) = 2p$; y puesto que A tiene el doble de B, $P(A) = 2P(B) = 4p$. Ahora como la suma de las probabilidades debe ser 1, tenemos:

$$
\begin{aligned}
p + 2p + 4p &= 1 \\
7p &= 1 \\
p &= \frac{1}{7}
\end{aligned}
$$

En consecuencia,

$$P(A) = 4p = \frac{4}{7}$$

$$P(B) = 2p = \frac{2}{7}$$

$$P(C) = p = \frac{1}{7}$$

Pregunta: ¿Cuál es la probabilidad de que B o C ganen?

$$
\begin{aligned}
&P(B,C) = P(B) + P(C) \\
&P(B,C) = \frac{2}{7} + \frac{1}{7} \\
&P(B,C) = \frac{3}{7}
\end{aligned}
$$

Ejemplo:

Se lanzan 2 monedas y se observa el número de caras que resulten. Calcular la probabilidad de que aparezca por lo menos una cara

Solución:

El espacio muestral es S = {0, 1, 2}. Se obtiene un espacio de probabilidad por medio de:

$$P(0) = \frac{1}{4},$$

$$P(1) = \frac{2}{4},$$

$$P(2) = \frac{1}{4}$$

Entonces la probabilidad de que aparezca por lo menos una cara es:

$$P(1,2) = P(1) + P(2) = \frac{2}{4} + \frac{1}{4} = \frac{3}{4} = 0{,}75 = 75\%$$

Ejemplo:

Se lanza solo un dado. Hallar la probabilidad de que resulte 2 ó 5.

Solución:

El espacio muéstrales δ = {l, 2, 3, 4, 5, 6}. Si se asignan probabilidades iguales a los puntos muéstrales, es decir si suponemos que el dado es honrado, entonces:

$$P(l) = P(2) = \ldots = P(6) = \frac{1}{6}$$

El suceso que resulte 2 ó 5. Por tanto

$$P(2\,ó\,5) = P(2) + P(5) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Probabilidad Condicional

Sean A y B dos sucesos (figura 18), tales que $P(A) > O$. La notación dada por $P(B \setminus A)$ significa la probabilidad de B dado que A ha ocurrido. Como se sabe que A ha ocurrido, se convierte en el nuevo espacio muestral remplazando el original δ. Asi se llega a la definición:

$$P(B \setminus \mathrm{A}) = \frac{P(A \cap B)}{P(A)}$$

$$P(B \setminus \mathrm{A}) * P(A) = P(A \cap B)$$

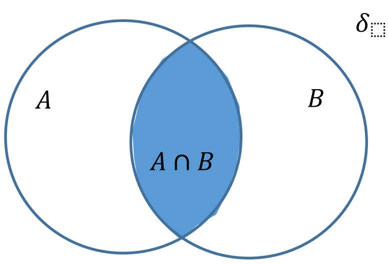

Figura 18. ***Probabilidad Condicional***
Fuente: Elaboración propia.

Ejemplo:

Hallar la probabilidad de que un sólo lanzamiento de un dado resulte en: a) un número menor que 4, b) que sea un número impar menor que 4.

Solución:

a. Si B denota el suceso {número en el dado menor que 4}. Ya que B es la unión de los sucesos 1, 2 ó 3 entonces la probabilidad:

$$P(B) = P(1) + P(2) + P(3) + P(4) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$$

Suponiendo probabilidades iguales para los elementos del espacio muéstrales.

b. Si A es el suceso {número en el dado impar menor que 4} observamos que $P(A) = 3/6 = 1/2$. También $P(A \text{ n } B) = 2/6 = 1/3$.

Entonces

$$P(B \setminus \mathrm{A}) = \frac{P(A \cap B)}{P(A)} = \frac{1/3}{1/2} = \frac{2}{3}$$

Por tanto, saber que el resultado del lanzamiento es un número impar aumenta la probabilidad de 1/2 a 2/3.

Sucesos Independientes

Dada la probabilidad, P(B\A) = P(B), quiere decir la probabilidad de que B ocurra y no tiene ninguna relación con la ocurrencia o no ocurrencia de A, esto implica que los dos eventos A y B son sucesos independientes. Esto es equivalente a

$$P(A \cup B) = P(A)\ P(B)$$

Como se deduce de

$$P(B \setminus \mathrm{A}) = \frac{P(A \cap B)}{P(A)}$$

$$P(B \setminus \mathrm{A}) * P(A) = P(A \cap B)$$

Inversamente, si se cumple $P(A \cup B) = P(A)\ P(B)$ entonces A y B son independientes. Algunas propiedades de la independencia están dadas en los problemas siguientes.

Si A lo A_1, A_2 A_3 son independientes entonces deben ser independientes por parejas,

$$P\left(A_j \cap A_k\right) = P\left(A_j\right)\ P\left(A_k\right) \qquad j \neq k \text{ donde } \quad j,\ k = 1, 2, 3$$

Y también debemos tener:

$$P\left(A_1 \cap A_2 \cap A_3\right) = P\left(A_1\right)\ P\left(A_2\right)\ P\left(A_3\right)$$

La generalización a más de tres sucesos se hace fácilmente.

Teorema o Regla de Bayes

Si los eventos $B_1, B_2, \dots, B_k$ constituyen una partición del espacio muestral S, $P(B_i) \neq 0$ para $i = 1, 2,\dots,k$, entonces, para cualquier evento A en S, tal que $P(A) \neq 0$,

$$P(B_r \backslash A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A \backslash B_r)}{\sum_{i=1}^{k} P(B_i)P(A \backslash B_i)}$$

Para $r = 1, 2, \dots, k$

Prueba:

Según la definición de probabilidad condicional,

$$P(B_r \backslash A) = \frac{P(B_r \cap A)}{P(A)}$$

Y después usando el teorema de probabilidad total en el denominador.

$$P(B_r \backslash A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A \backslash B_r)}{\sum_{i=1}^{k} P(B_i)P(A \backslash B_i)}$$

Que completa la demostración.

Teorema de probabilidad total

Si los eventos B_1, B_2, … B_k constituyen una partición del espacio muestral S, tal que $P(B_i) \neq 0$ para i= 1, 2, …k, entonces, para cualquier evento A de S,

$$P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} P(B_i)\, P(A \backslash B_i)$$

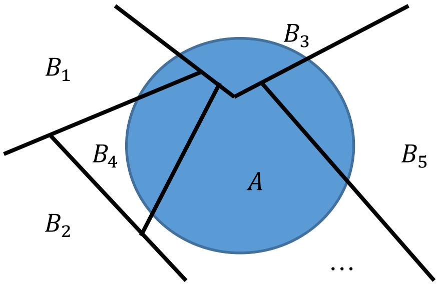

Figura 19. ***Partición del espacio muestral S***
Fuente: Elaboración propia.

Ejemplo:

Una empresa de manufactura emplea tres planos analíticos para el diseño y desarrollo de un producto específico. Por razones de costos los tres se utilizan en momentos diferentes. De hecho, los planos 1, 2 y 3 se utilizan para 30%, 20% y 50% de los productos, respectivamente. La tasa de defectos difiere en los tres procedimientos de la siguiente manera,

$$P(D \mid P1) = 0.01,\ P(D \mid P2) = 0.03,\ P(D \mid P3) = 0.02,$$

En donde $P(D \mid Pj)$ es la probabilidad de que un producto esté defectuoso, dado el plano *j*. Si se observa un producto al azar y se descubre que está defectuoso, ¿cuál de los planos tiene más probabilidades de haberse utilizado y, por lo tanto, de ser el responsable?

Solución:

A partir del planteamiento del problema

$P(P1) = 0.30$, $P(P2) = 0.20$ y $P(P3) = 0.50$,

Debemos calcular $P(Pj \mid D)$ para j = 1, 2, 3. La regla de Bayes muestra

$$P(P_1 \backslash D) = \frac{P(P_1)\ P(D \backslash P_1)}{P(P_1)\ P(D \backslash P_1) + P(P_2)\ P(D \backslash P_2) + P(P_3)\ P(D \backslash P_3)}$$
$$= \frac{(0.30)\ (0.01)}{(0.30)\ (0.01) + (0.20)\ (0.03) + (0.50)\ (0.02)}$$
$$= \frac{0.003}{0.019} = 0.158$$

De igual manera,

$$P(P_2 \backslash D) = \frac{P(P_2)\ P(D \backslash P_2)}{P(P_1)\ P(D \backslash P_1) + P(P_2)\ P(D \backslash P_2) + P(P_3)\ P(D \backslash P_3)}$$
$$= \frac{(0.30)\ (0.20)}{(0.30)\ (0.01) + (0.20)\ (0.03) + (0.50)\ (0.02)}$$
$$= 0.316$$

$$(P_3 \backslash D) = \frac{P(P_3)\ P(D \backslash P_3)}{P(P_1)\ P(D \backslash P_1) + P(P_2)\ P(D \backslash P_2) + P(P_3)\ P(D \backslash P_3)}$$
$$= \frac{(0.50)\ (0.02)}{(0.30)\ (0.01) + (0.20)\ (0.03) + (0.50)\ (0.02)}$$
$$= 0.526$$

Se pude observar en los resultados anteriores que la probabilidad condicional de ser defectuoso, dado el plano 3, es mayor que con los planos 1 y 2; por lo tanto, un defecto en un producto elegido al azar tiene más posibilidad de ser el resultado de haber usado el plano 3.

Calculo de Probabilidades

Ejemplo:

En un juego de cartas se extrae aleatoriamente una, de una baraja de 52 cartas. Encontrar la probabilidad de sacar *(a)* una carta de as, *(b)* una carta con jota de corazones, *(c)* una carta con tres de tréboles o un seis de diamantes, *(d)* una cara de corazón, *(e)* cualquier carta excepto corazones, (f) una cara de diez o una pica, (g) una carta que no sea ni un cuatro ni un trébol.

Solución:

Por facilidad utilicemos C, *P, D*, *T* para indicar los siguientes palos: corazón, pica, diamante y trébol, respectivamente, y 1, 2..., 13 por as, dos..., rey 3 $\cap$ C significa tres de corazones, en tanto que 3 U C significa tres o corazón. Empleemos el espacio muestral y asignando probabilidades iguales de *1/52* a cada punto muestral. Así, por ejemplo,

a.

$$
\begin{aligned}
P(1) &= P(1\cap C \text{ ó } 1\cap P \text{ ó} 1\cap D \text{ ó } 1\cap T) \\
&= P(1\cap C) + P(1\cap P) + P(1\cap D) + P(1\cap T) \\
&= \frac{1}{52}+\frac{1}{52}+\frac{1}{52}+\frac{1}{52}=\frac{1}{13}
\end{aligned}
$$

b. $P(11\cap C) = \dfrac{1}{52}$

c. $P(3\cap T \text{ ó } 6\cap D) = P(3\cap T) + P(6\cap D) = \dfrac{1}{52}+\dfrac{1}{52}=\dfrac{1}{26}$

d. $P(1\cap C \text{ ó } 2\cap C \text{ ó} \ldots 13\cap C) = \dfrac{1}{52}+\dfrac{1}{52}+\ldots+\dfrac{1}{52}=\dfrac{13}{52}=\dfrac{1}{4}$

También se había podido llegar a este resultado observando que hay cuatro palos y al ceder uno tiene una probabilidad igual de ser extraído, esto es 1/4.

e. *(e)* $P(C') = 1 - P(C) = 1 - 1/4 = 3/4$

f. *(f)* Puesto que 10 y *P* no son mutuamente excluyentes tenemos del Teorema

$$P(10 \cup P) = P(10) + P(P) - P(10 \cap P) = 1/13 + 1/4 - 1/52 = 4/13$$

g. La probabilidad de no ser cuatro o trébol puede denotarse por $P(4' \cap T')$. Pero por el Teorema, $4' \cap T' = (4 \cup T)'$. Por tanto

$$P(4' \cup T') = P(4 \cup T)' = 1 - P(4 \cup T) = 1 - [P(4) + P(T) - P(4 \cap T)]$$
$$= 1 - \left[\frac{1}{13} + \frac{1}{4} - \frac{1}{52}\right] = \frac{9}{13}$$

Ejemplo:

Una bola se extrae aleatoriamente de una caja que contiene 6 bolas rojas, 4 bolas blancas y 5 bolas azules. Determinar la probabilidad de que sea *(a)* roja, *(b)* blanca, *(c)* azul, *(d)* no roja, *(e)* roja o blanca.

Solución:

Método l

Denótese por *n, B* y *A* los sucesos de extraer una bola roja, blanca y azul, respectivamente. Entonces

$$P(R)=\frac{maneras\ \ de\ \ eligir\ \ una\ \ bola\ \ roja}{maneras\ \ de\ \ eligir\ \ una\ \ bola}=\frac{6}{6+4+5}=\frac{6}{15}=\frac{2}{5}$$

Método 2

Nuestro espacio muestral consiste de 6 + 4 + 5 = 15 puntos muéstrales. Entonces si asignamos probabilidades iguales 1/15 a cada punto muestral observamos que *P(R)* = 6/15 = 2/5, debido a que hay 6 puntos muéstrales que' corresponden a "bola roja",

a. $P(B)=\frac{4}{6+4+5}=\frac{4}{15}$

b. $P(B)=\frac{5}{6+4+5}=\frac{5}{15}=\frac{1}{3}$

c. $P(no\ \ roja)=P(R')=1-P(R)=1-\frac{2}{5}=\frac{3}{5}$ por parte (a)

d. Método 1.

$$P(roja\ \ o\ \ blanca)=P(R\cup B)=\frac{maneras\ \ de\ \ eligir\ \ una\ \ bola\ \ roja\ \ o\ \ blanca}{maneras\ \ de\ \ eligir\ \ una\ \ bola}$$

$$=\frac{6+4}{6+4+5}=\frac{10}{15}=\frac{2}{3}$$

También puede resolverse utilizando el espacio muestral como en la parte *(a)*.

Método 2

$$P(R\cup B)=P(A')=1-P(A)=1-\frac{1}{3}=\frac{2}{3}\ \text{Por parte } (c)$$

Probabilidad condicional y sucesos independientes

Ejemplo:

En un juego de casino se tiene un dado no cargado, y se lanza dos veces. Hallar la posibilidad de sacar un 4, 5 ó 6 en el primer lanzamiento y un 1, 2, 3 ó 4 en el segundo lanzamiento.

Solución:

Sean A_1 el suceso "4, 5 ó 6 en el primer lanzamiento" y A^2 el suceso "1, 2, 3 ó 4 en el segundo lanzamiento". Luego estamos buscando $P(A_1 \cap A_2)$

Método l

$$P(A_1 \cap A_2) = P(A_1)P(A_2 / A_1) = P(A_1)P(A_2) = \left(\frac{3}{6}\right)\left(\frac{4}{6}\right) = \frac{1}{3}$$

Hemos empleado aquí el hecho de que el resultado del segundo lanzamiento es *independiente* del primero así que $P(A_2 \backslash A_1) = P(A_2)$. También hemos usado $P(A_1.) = 3/6$ (ya que 4, 5 ó 6 son 3 resultados de las 6 probabilidades igualmente factibles) y $P(A_2) = 4/6$ (ya que 1,2, 3 ó 4 son 4 resultados de las 6 probabilidades igualmente factibles).

Método 2

Para cada una de las 6 formas en las cuales el dado cae, desde que se realizó el primer lanzamiento, puede asociarse con cada una de las 6 maneras en que cae el dado una vez realizado el segundo lanzamiento, un total de 6.6 = 36 maneras, todas igualmente posibles.

Las tres formas en que A_1 ocurre se puede asociar con cada una de las 3 formas en que A_2 ocurre para dar 3.4 = 12 maneras en que tanto A_1 como A_2 ocurren. Entonces

$$P(A_1 \cap A_2) = \frac{12}{36} = \frac{1}{3}$$

Esto indica directamente que A_1 y A_2 son independientes puesto que

$$P(A_1 \cap A_2) = \left(\frac{3}{6}\right)\left(\frac{4}{6}\right) = \frac{1}{3} = P(A_1)\ P(A_1)$$

Ejemplo:

En una partida de cartas se sacan dos cartas de un juego de 52 barajas. Encuentre la probabilidad de que ambas sean unos o ases si la carta *(a)* con sustitución, *(b)* sin sustitución.

Método l

Consideremos A_1 como suceso "as en la primera carta sacada" y A_2 el suceso "as en la segunda carta extraída". Estamos buscando $P(A_1 \cap A_2) = P(A_1)P(A_2 / A_1)$

a. Para la primera extracción de la carta hay 4 ases en las 52 barajas, $P(A)$ = 4/52. Por otro lado, si la carta se sustituye para la segunda extracción, se tiene que $P(A_2/A_1)$ = 4/52, ya que se cuenta con 4 ases en las 52 barajas para la segunda extracción. Por lo tanto:

$$P(A_1 \cap A_2) = P(A_1)P(A_2 / A_1) = \left(\frac{4}{52}\right)\left(\frac{4}{52}\right) = \frac{1}{169}$$

b. De la parte (a), $P(A_1)$ = 4/52, si ocurre una carta con un as en la primera extracción tendremos 3 en las 51 cartas restantes, así que $P(A_2/A_1)$ = 3/51. Entonces

$$P(A_1 \cap A_2) = P(A_1)P(A_2 / A_1) = \left(\frac{4}{52}\right)\left(\frac{3}{51}\right) = \frac{1}{221}$$

Método 2

a. Se puede apreciar que la primera carta se puede sacar de una de las 52 opciones posibles y como existe reemplazamiento para sacar la segunda baraja, puede extraerse también de una de las 52 formas posibles. Luego se pueden sacar ambas cartas de (52)(52) maneras, todas potencialmente posibles.

Como hay 4 opciones de obtener una carta con un as en la primera extracción y 4 maneras de obtener una carta con un as en la segunda extracción entonces las maneras de obtener ases en la primera y segunda extracción es (4)(4). Así la posibilidad es:

$$\frac{(4)(4)}{(52)(52)} = \frac{1}{169}$$

b. Como la primera baraja puede obtenerse de una de las 52 opciones posibles y no existe reemplazamiento en la segunda baraja, se puede obtener de una de las 51 maneras posibles. Luego ambas barajas se pueden extraer de (52)(51) maneras, todas igualmente posibles. Como hay 4 maneras de sacar una baraja con un as en la primera extracción y 3 maneras de sacar una carta con un as en la segunda extracción, esto implica que el número de maneras de sacar una carta con un as en la primera y segunda extracción es (4)(3). Así la posibilidad es:

$$\frac{(4)(3)}{(52)(51)} = \frac{1}{221}$$

Ejemplo:

En un sorteo se procede a extraer tres bolas continuamente de una urna que contiene 6 bolas de color rojo, 4 bolas de color blanco y 5 bolas de color azul. Calcule la posibilidad de sacar en el orden roja, blanca y azul si las bolas *(a)* se sustituyen o remplazan, *(b)* no se sustituyen o remplazan.

Solución:

Si R_1 = suceso "roja en la primera extracción", B_2 = suceso "blanca en la segunda extracción" y A_3 = suceso "azul en la tercera extracción". Requerimos $P(R_1 \cap B_2 \cap A_3)$.

a. Si cada bola se reemplaza, entonces los sucesos son independientes y

$$P(R_1 \cap B_2 \cap A_3) = P(R_1)\ P(B_2/R_1)\ P(A_3/R_1 \cap B_2)$$
$$= P(R_1)\ P(B_2)\ P(A_3)$$
$$= \left(\frac{6}{6+4+5}\right)\left(\frac{4}{6+4+5}\right)\left(\frac{5}{6+3+5}\right) = \frac{8}{225}$$

b. Si no se remplazan las bolas, entonces los sucesos son dependientes y

$$P(R_1 \cap B_2 \cap A_3) = P(R_1)\ P(B_2/R_1)\ P(A_3/R_1 \cap B_2)$$
$$= P(R_1)\ P(B_2)\ P(A_3)$$
$$= \left(\frac{6}{6+4+5}\right)\left(\frac{4}{5+4+5}\right)\left(\frac{5}{5+3+5}\right) = \frac{4}{91}$$

Ejemplo:

Hallar la probabilidad de obtener al menos un 4 en dos lanzamientos de un dado honrado.

Solución

Sea A_1 = suceso "4 en el primer lanzamiento" y A_2 = suceso "4 en el segundo lanzamiento", Así $A_1 \cup A_2$ = suceso "4 en el primer lanzamiento o 4 en el segundo lanzamiento o ambos" = suceso "al menos un 4"

Requeridos $P(A_1 \cup A_2)$

Método l

Los sucesos A_1 y A_2 no son mutuamente excluyentes, pero son independientes. Por tanto,

$$
\begin{aligned}
&P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)\\
&= P(A_1) + P(A_2) - P(A_1)P(A_2)\\
&= \frac{1}{6} + \frac{1}{6} - \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{11}{36}
\end{aligned}
$$

Método 2

*P(*Sacar al menos un 4) + *P* (No sacar un 4) = 1

De aquí se obtiene que

$$
\begin{aligned}
&P(al\ menos\ un\ 4) = 1 - P(ningún\ 4)\\
&= 1 - P(no\ \ 4\ en\ el\ \ 1er.\ \ lanzamiento\ y\ no\ \ 4\ en\ el\ \ 2^{o}.lanzamiento)\\
&= 1 - P(A_1' \cap A_2')\\
&= 1 - P(A_1')\,P(A_2')\\
&= 1 - \left(\frac{5}{6}\right)\left(\frac{5}{6}\right) = \frac{11}{36}
\end{aligned}
$$

Método 3

Número total de maneras igualmente factibles en las que ambos dados pueden caer = 6 .6 = 36.

También,

Número de formas en que A_1 ocurra pero no A_2 = 5.
Número de formas en que A_2 ocurra pero no A_1 = 5.
Número de formas en que A_1 y A_2 ocurran = 1

Por lo tanto, el número de maneras en las que por lo menos uno de los sucesos como es A_1 ó A_2 suceda es = 5 + 5 + 1 = 11. Lo que implica que $P(A_1 \cup A_2)$ = 11/36.

Ejemplo:

Una urna X contiene 4 bolas de color blanco y 2 bolas de color negro; otra urna Z contiene 3 bolas de color blanco y 5 bolas de color negro. Si se extrae una bola de cada urna, hallar la posibilidad de que *(a)* las dos sean de color blanco, *(b)* las dos sean de color negro, *(c)* una sea blanca y la otra sea negra.

Solución:

Llamemos a B_1 = el suceso "sacar una bola blanca de la primera urna", B_2 = el suceso "sacar una bola blanca de la segunda urna".

(a) $P(B_1 \cap B_2) = P(B_1)P(B_2 / B_1) = P(B_1)P(B_2) = \left(\frac{4}{4+2}\right)\left(\frac{3}{3+5}\right) = \frac{1}{4}$

(b) $P(B_1' \cap B_2') = P(B_1')P(B_2' / B_1') = P(B_1')P(B_2') = \left(\frac{2}{4+2}\right)\left(\frac{5}{3+5}\right) = \frac{5}{24}$

(c) La probabilidad pedida es

$$1 - P(B_1 \cap B_2) - P(B_1' \cap B_2') = 1 - \frac{1}{4} - \frac{5}{24} = \frac{13}{24}$$

Ejemplo:

La cesta I contiene 3 bolas de color rojo y 2 de color azul, mientras que la cesta II contiene 2 bolas de color rojo y 8 de color azul. Si se lanza una moneda no cargada y se obtiene cara, se saca una bola de la cesta I*;* si se obtiene sello se saca una bola de la cesta II. Calcule la posibilidad de sacar una bola de color roja.

Solución:

Sea R el suceso de "sacar una bola de color roja" mientras que I y II indican los sucesos de escoger cesta I y cesta II, en ese orden. Luego, la posibilidad de sacar una bola de color roja es:

$$P(R) = P(I)P(R/I) + P(II)P(R/II) = \left(\frac{1}{2}\right)\left(\frac{3}{3+2}\right) + \left(\frac{1}{2}\right)\left(\frac{2}{2+8}\right) = \frac{2}{5}$$

Probabilidad utilizando Análisis Combinatorio

Ejemplo:

Una caja contiene 8 bolas rojas, 3 blancas y 9 azules. Si se extrae 3 bolas aleatoriamente sin remplazar, determinar la probabilidad de que *(a)* las 3 bolas sean rojas, *(b)* las 3 bolas sean blancas, *(c)* 2 sean rojas y 1 blanca, *(d)* al menos 1 sea blanca, *(e)* se extraiga una de cada color y (f) las bolas sean extraídas en el orden rojo, blanco, azul.

Solución:

Método 1

Sean R_1, R_2, R_3 los sucesos, "bola roja en la primera extracción", "bola roja en la segunda extracción", "bola roja en la tercera extracci6n", respectivamente. Así, $R_1 \cap R_2 \cap R_3$ representa el suceso "las 3 bolas extraídas son rojas". De esta manera tenemos:

$$P(R_1 \cap R_2 \cap R_3) = P(R_1)P(R_2 / R_1)P(R_3 / R_1 \cap R_2)$$
$$= \left(\frac{8}{20}\right)\left(\frac{7}{19}\right)\left(\frac{6}{18}\right) = \frac{14}{285}$$

Método 2

a. Probabilidad perdida es:

$$\frac{\textit{número de grupos de 3 bolas entre 8 rojas}}{\textit{número de grupos de 3 bolas entre 20}} = \frac{{}_8C_3}{{}_{20}C_3} = \frac{14}{285}$$

b. Empleando el segundo método indicado en la parte (a):

$$P(3\, bolas\, blancas) = \frac{{}_3C_3}{{}_{20}C_3} = \frac{1}{1140}$$

También se puede utilizar el primer método indicado en la parte *(a)*.

c. 2 sean rojas y 1 blanca:

$$P(2\, bolas\, rojas\ y\ 1\, blancas) = \frac{\text{(grupos de 2 entre 8 bolas rojas)(grupos de 1 entre 3 bolas blancas)}}{\text{número de grupos de 3 bolas entre 20}}$$
$$= \frac{\left({}_8C_2\right)*\left({}_3C_1\right)}{{}_{20}C_3} = \frac{7}{95}$$

d. Al menos 1 sea blanca:

$$P(ninguna\, blanca) = \frac{{}_{17}C_3}{{}_{20}C_3} = \frac{34}{57}$$

Entonces, $P(almenos\ 1 blanca) = 1 - \frac{34}{57} = \frac{23}{57}$

e. Se extraiga una de cada color.

$$P(1\ de\ cada\ color)=\frac{\left({}_{8}C_{1}\right)*\left({}_{3}C_{1}\right)*\left({}_{9}C_{1}\right)}{{}_{20}C_{3}}=\frac{18}{95}$$

f. Las bolas se sacan en orden rojo, blanco, azul.

$$P(extraer\ \ las\ \ bolas\ \ en\ \ orden\ \ rojo, blanco, azul)=\frac{1}{3!}P(1\ \ de\ \ cada\ \ color)$$

$$=\frac{1}{6}\left(\frac{18}{95}\right)=\frac{3}{95}$$

Ejemplo:

De una partida de cartas se sacan 5 cartas de un paquete de 52 cartas. Calcule la posibilidad de extraer *(a)* 4 cartas con ases, *(b)* 4 cartas con ases y un rey, *(e)* 3 cartas con dieces y 2 jotas, *(d)* las cartas con números 9, 10, jota, reina, rey en cualquier orden, *(e)* 3 carta de un palo y 2 de otro, (f) al menos una carta con un as.

Solución:

a. $$P(4\ ases)=\frac{\left({}_{4}C_{4}\right)\left({}_{48}C_{1}\right)}{{}_{52}C_{5}}=\frac{1}{54145}$$

b. $$P(4\ ases\ y\ un\ rey)=\frac{\left({}_{4}C_{4}\right)\left({}_{4}C_{1}\right)}{{}_{52}C_{5}}=\frac{1}{649740}$$

c. $$P(3\ dieces\ y\ 2\ jotas)=\frac{\left({}_{4}C_{3}\right)\left({}_{4}C_{2}\right)}{{}_{52}C_{5}}=\frac{1}{108290}$$

d. $$P(\text{nueve, diez, jota, reina, rey})=\frac{\left({}_{4}C_{1}\right)\left({}_{4}C_{1}\right)\left({}_{4}C_{1}\right)\left({}_{4}C_{1}\right)\left({}_{4}C_{1}\right)}{{}_{52}C_{5}}=\frac{64}{162435}$$

e. $$P(3\text{ de un palo, }2\text{ de otro})=\frac{\left(4*{}_{13}C_{3}\right)\left(3*{}_{13}C_{2}\right)}{{}_{52}C_{5}}=\frac{429}{4165}$$

Como hay 4 formas de seleccionar el primer palo y 3 formas de escoger el segundo

f. $$P(ningun\ ases) = \frac{_{48}C_5}{_{52}C_5} = \frac{35673}{54145}$$

Luego,

$$P(almenos\ un\ ases) = 1 - \frac{35673}{54145} = \frac{18472}{54145}$$

Ejemplo:

Si un dado es lanzado cinco veces, calcule la posibilidad de sacar (3) seises.

Solución:

Si se representan los lanzamientos, cinco lanzamientos del dado como: - - - -. Cada espacio tendrá los sucesos 6 o no (6'). Al analizar que tres 6 y dos no (6') pueden ocurrir que salga como: 666' 6 6' ó 66' 6 6' 6, etc.

Así la probabilidad del resultado 6 6 6' 6 6' es

$$P(666'66') = P(6)P(6)P(6')P(6)P(6') = \frac{1}{6}\frac{1}{6}\frac{5}{6}\frac{1}{6}\frac{5}{6} = \left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$$

Al suponer que los sucesos son independientes. Análogamente:

$$P = \left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$$

Los otros resultados en los cuales se saquen tres 6 y dos no (‘6). Tenemos entonces $_5C_3$ = 10 de estos sucesos que se caracterizan por ser excluyentes. Por lo tanto, la posibilidad pedida es

$$P(666'66' \ o \ 66'66'6 \ o \ldots) = {}_5C_3\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 = \frac{5!}{3! \ 2!}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 = \frac{125}{3888}$$

En general, si $p = P(A)$ y $q = 1 - p = P(A')$, por el mismo razonamiento anterior la probabilidad de obtener exactamente x veces A en n ensayos independientes es

$$_nC_x p^x q^{n-x} = \binom{n}{x} p^x q^{n-x}$$

Ejemplo:

Un armario contiene 6 textos de matemáticas y 4 textos de física. Encuentre la probabilidad de que 3 textos de matemáticas estén juntos.

Solución:

Los textos pueden ordenarse entre sí de $_{10}P_{10}$ = 10! formas. Si se supone que los 3 libros determinados de matemáticas se remplazan por 1. Así tenemos un total de 8 libros que pueden ordenarse entre sí de $_8P_8$ = 8! formas. Pero los 3 libros de matemática pueden ordenarse entre sí de $_3P_3$ = 3! formas. La probabilidad está dada por:

$$\frac{8! \ 3!}{10!}$$

AUTOEVALUACIÓN

PROBLEMAS SOBRE PERMUTACIONES

1. Encuentre el número de permutaciones de 6 elementos, conformados por las letras *a, b, c, d, e, f*, si se toman 4 a la vez. Es decir, hallar el número de ¨palabras de 4 letras diferentes¨ que pueden armarse con las seis letras mencionadas.
2. ¿Cuántas permutaciones de tres elementos se forman con tres objetos *a, b, c*?
3. Suponer que una placa de automóvil consta de dos letras distintas seguida de tres dígitos de los cuales el primero no es cero. ¿Cuántas placas diferentes pueden grabarse?
4. ¿Calcule el número de señales diferentes que pueden armarse al colocar 8 banderas en línea, de un conjunto de 4 banderas de color rojo, 3 banderas de color blanco y una azul?
5. Si no se permiten repeticiones, (i) ¿cuántos números de 3 dígitos se pueden formar con los seis dígitos 2, 3, 5, 6, 7 y 9? (ii) ¿Cuántos son menores que 500? (iii) ¿Cuántos son números pares? (iv) ¿Cuántos son números impares? (v) ¿Cuántos son múltiplo de 3?
6. ¿De cuántas formas se puede sentar en una reunión siete personas, (i) si se sientan en fila de siete sillas, (ii) si se sientan alrededor de una mesa redonda?
7. (i) ¿Si a una reunión llegan 3 niños y 2 niñas, de cuantas maneras pueden sentarse en una fila? (ii) ¿De cuántas maneras se podrían sentar si los niños se sientan juntos y las niñas también? (iii) ¿De cuántas maneras se podrían sentar en fila si las niñas se sientan juntas?

8. ¿Calcular el número de señales diferentes que podrían armarse con 6 banderas colgadas en una línea, de un conjunto de 4 banderas de color rojo y 2 de color azul?
9. ¿Hallar el número de permutaciones distintas que pueden formarse con todas las letras de cada una de las siguientes palabras: (i) tema, (ii) campana, (iii) estadísticas?
10. ¿De cuántas maneras 3 americanos, 4 franceses, 4 colombianos y 2 italianos pueden sentarse en una fila, de modo que los de la misma nacionalidad se sienten juntos?
11. Suponer que una urna contiene 8 bolas. Hallar el número de pruebas ordenadas de tamaño 3. (i) Con sustitución y (ii) sin sustitución.
12. (i) ¿Cuántas placas para automóvil se pueden hacer, si cada placa consta de dos letras diferentes seguidas de 3 dígitos diferentes? (ii) Resolver el problema, si el primer dígito no puede ser 0, 5, 7 o 9.
13. Hallar el número de maneras en que 6 personas pueden conducir un tobogán (especie de trineo), si uno de tres debe manejar.
14. (i) Hallar el número de palabras de 4 letras que pueden formarse con las letras de la palabra CRISTAL. (ii) ¿Cuáles de ellas contienen sólo consonantes? (iii) ¿Cuántas empiezan y terminan por consonantes? (iv) ¿Cuántas empiezan por vocal? (v) ¿Cuántas contienen la letra T? (vi) ¿Cuántas de estas empiezan con la letra T y terminan por vocal? (vii) ¿Cuántas empiezan por T y también contienen S? (viii) ¿Cuántas contienen ambas vocales?

PROBLEMAS SOBRE COMBINACIONES

1. ¿De cuántas maneras puede escogerse un comité, compuesto de 3 hombres y 2 mujeres, de un grupo de 7 hombres y 5 mujeres?
2. Un comité de 5 representantes estudiantiles del colegio JC son seleccionados cada año. (i) ¿De cuántas formas pueden elegir el comité si hay 15 estudiantes elegibles? (ii) ¿De cuántas maneras, si dos de los estudiantes elegibles no asisten simultáneamente? (iii) ¿De cuántas maneras si dos de los estudiantes elegibles son novios y solo asisten juntos?
3. En un examen de cálculo, el estudiante tiene que contestar 7 de 10 preguntas. (i) ¿Cuántas formas tiene de escoger? (ii) ¿Cuántas formas tiene que escoger, si las 4 primeras preguntas son obligatorias? (iii) ¿Cuántas, si tiene que contestar 4 de las 5 primeras preguntas?
4. Hallar el número de subconjuntos de un conjunto X que contiene n elementos.
5. ¿De Cuántas maneras puede un profesor escoger uno o más estudiantes de seis elegibles?
6. Calcular

$$(i)\binom{5}{2}, \quad (ii)\binom{7}{3} \quad (ii)\binom{14}{2} \quad (iii)\binom{6}{4} \quad (iv)\binom{20}{17} \quad (v)\binom{18}{15}$$

7. Calcular

$$(i)\binom{9}{3,\ 5,\ 1} \quad (ii)\binom{7}{3,\ 2,\ 2,\ 0} \quad (iii)\binom{6}{2,\ 2,\ 1,\ 0}.$$

8. Un curso de aritmética tiene 10 niños y 4 niñas. (i) ¿De cuántas formas el profesor puede escoger un comité de cinco? (ii) ¿Cuántos comités constarán con una niña por lo menos? (iii) ¿Cuántos tendrán solo una niña?
9. Un restaurante tiene 11 clientes de confianza. (i) ¿De cuántas maneras puede invitar 6 de ellos a cenar? (ii) ¿De cuántas maneras, si dos son casados y asisten juntos? (iii) ¿De cuántas maneras si dos de ellos son enemigos y no asisten simultáneamente?
10. Una prueba de suficiencia tiene 15 preguntas, los usuarios solo deben contestar 10 preguntas. (i) ¿Cuántas maneras tiene de escoger? (ii) ¿Cuántas, si las tres de primeras son obligatorias? (iii) ¿Cuántas, si dos de las cuatro primeras son obligatorias? (iv) ¿Cuántas, si tiene que contestar exactamente 3 de las 6 primeras? (v) ¿Cuántas, si tiene que contestar por lo menos 4 de las 5 primeras?
11. En un juego de "póker" se tienen que repartir 5 cartas. ¿De cuántas formas puede repartirlas si se pide: (i) una escalera flor, (ii) un "póker", (iii) una escalera, (iv) un par de ases y (v) un par cualquiera (dos cartas iguales)?

PROBLEMAS DE INTRODUCCIÓN A LA PROBABILIDAD

1. Al lanzar 3 monedas, observar el número de caras que resulten.
2. Sea A el evento en que aparecen una cara por lo menos y sea B el evento en que aparecen todas las caras o todos los sellos.

3. Se supone un espacio muestral S que consta de 4 elementos: S = {a_1, a_2, a_3, a_4}. ¿Qué función define un espacio de probabilidad S?

   a. $P(a_1)=\frac{1}{2},\ P(a_2)=\frac{1}{3},\ P(a_3)=\frac{1}{4},\ P(a_4)=\frac{1}{5}$
   b. $P(a_1)=\frac{1}{2},\ P(a_2)=\frac{1}{4},\ P(a_3)=-\frac{1}{4},\ P(a_4)=\frac{1}{2}$
   c. $P(a_1)=\frac{1}{2},\ P(a_2)=\frac{1}{3},\ P(a_3)=\frac{1}{8},\ P(a_4)=\frac{1}{8}$
   d. $P(a_1)=\frac{1}{2},\ P(a_2)=\frac{1}{4},\ P(a_3)=\frac{1}{4},\ P(a_4)=0$

4. Sea S = {a_1, a_2, a_3, a_4}, y sea P una función de probabilidad de S.

   a. Hallar $P(a_1)$ si $P(a_2)=\frac{1}{3},\ P(a_3)=\frac{1}{6},\ P(a_4)=\frac{1}{9}$
   b. Hallar $P(a_1)$ y $P(a_2)$ si $P(a_3)=P(a_4)=\frac{1}{4}$ y $P(a_1)=2P(a_2)$
   c. Hallar $P(a_1)$ si $P(a_2,a_3)=\frac{2}{3},\ P(a_2,a_4)=\frac{1}{2}$ y $P(a_2)=\frac{1}{3}$

5. Una moneda está cargada (aumentada de peso) de modo que la posibilidad de salir cara (C), sea el doble de la de sello (S). Hallar $P(C)$ y $P(S)$

6. En un torneo de ajedrez intervienen cuatro caballeros y seis damas. Los del mismo sexo tienen iguales probabilidades de ganar, pero cada caballero tiene el doble de la posibilidad de ganar que una dama. (i) Hallar la probabilidad de que gane una dama el torneo. (ii) Si hay una pareja y son casados, hallar la probabilidad de que uno de ellos gane el torneo.

7. En un campeonato de natación, tres estudiantes *A, B* y *C* participan. *A* y *B* tienen la misma probabilidad de ganar y el triple de la de *C*. Hallar la probabilidad de que gane A, *B* o *C.*
8. Un dado se arregla de manera que los números pares tienen el doble de posibilidad de salir que los impares. Calcule la posibilidad de que, (i) aparezca un número par, (ii) aparezca un número impar y (iii) aparezca un número primo impar.

PROBLEMAS DE PROBABILIDAD CON ANÁLISIS COMBINATORIO

1. Un comité estudiantil está conformado por 3 estudiantes de primero, 5 de segundo, 7 de penúltimo y 8 de último año. Se escoge un estudiante al azar para representar el comité. Hallar la probabilidad de que el estudiante sea, (i) de segundo, (ii) de último año o (iii) de penúltimo año.
2. En una partida de cartas se escoge una carta al azar entre 52 cartas numeradas del 1 a 52. Hallar la posibilidad de que el número de la carta sea, (i) divisible por 3, (ii) primo o (iii) sea par.
3. Se sacan dos cartas al azar de una baraja de 52 cartas. Hallar la probabilidad de que: (i) las dos sean espada y (ii) una sea espada y la otra corazón.
4. Se escogen al azar 3 lámparas entre 15 de una tienda, de las cuales 5 se encuentran defectuosas. Hallar las probabilidades de que: (i) ninguna sea defectuosa, (ii) una exactamente sea defectuosa y (iii) una por lo menos este defectuosa.

5. En un juego de naipes se toman dos cartas entre 13 cartas numeradas de 1 a 13. Hallar la posibilidad de que la suma sea impar, si (i) las dos cartas se sacan juntas o (ii) se saca una sin sustitución.
6. Un curso cuenta con 15 niñas, 3 tienen ojos de color azul. Si se seleccionan tres niñas al azar, ¿hallar la probabilidad de que, (i) las tres tengan ojos de color azul, (ii) ninguna tenga ojos de color azul, (iii) una por lo menos tenga ojos de color azul?
7. En un almacén de respuestas en una gaveta se encuentran cinco tornillos y cinco tuercas. Si se escogen tres piezas al azar, ¿cuál es la posibilidad de sacar un tornillo y dos tuercas?
8. Un curso tiene veinte estudiantes, A, B, C... Si se escoge un comité de 5, al azar, hallar la probabilidad de que (i) A pertenezca al comité, (ii) B y C pertenezcan al comité, (iii) A, B y C pertenezcan al comité y (iv) A o B pertenezcan al comité.
9. Un equipo consta de 8 niñas y 12 niños. Si se escoge al azar un comité de 6, hallar la posibilidad de (i) seleccionar cuatro niños, (ii) seleccionar exactamente 3 niños, (iii) seleccionar por lo menos un niño y (iv) seleccionar exactamente cinco niñas.
10. Un par de dados no arreglados se lanzan. Calcular la probabilidad de que la suma de los dos números sea mayor que 5.
11. En una institución educativa de 150 estudiantes, 70 estudian ingeniería, 60 estudian derecho, y 30 ingeniería y derecho. Si se selecciona un estudiante al azar, hallar la posibilidad de que el estudiante (i) estudie ingeniería y derecho y (ii) no estudie derecho ni ingeniería.

12. En un tren (5) niños y (4) niñas se sientan en fila. Hallar la probabilidad de que, (i) las niñas se sienten juntas y (ii)los niños y las niñas se sienten alternados.

PROBLEMAS DE PROBABILIDAD CON PROCESOS ESTOCÁSTICOS Y DIAGRAMAS DE ÁRBOL

1. Si se toman tres urnas con las siguientes condiciones:

Urna I:	Contiene 15 lámparas de las cuales 4 son defectuosas.
Urna II:	Contiene 8 con 2 defectuosas.
Urna III:	Contiene 10 con 3 defectuosas.

Si se selecciona al azar una urna y luego se saca una lámpara. Calcular la probabilidad *p* de que la lámpara no sea defectuosa.

2. Se lanza una moneda arreglada de modo que P(C) = 3/5 y P(S) = 2/5. Si cae cara, se selecciona al azar un número del 1 al 9 y si cae sello se selecciona al azar un número del 1 al 6. Hallar la posibilidad *p* de que se escoja un número impar.
3. Una fábrica cuenta con tres máquinas A, B y C que tienen una producción de 50%, 35% y 15% respectivamente del número de artículos que elaboran las fábricas. Los porcentajes de artículos desperfectos en la producción de estas máquinas son 4%, 5% y 6%. Si se escoge un articula al azar. Calcule la probabilidad de que el artículo no sea defectuoso.
4. En un bolso hay 8 bolas de color rojo y 4 bolas de color blanco. Se sacan 4 bolas del bolso, una tras otra. ¿Cuál es la probabilidad *p* de que las tres primeras sean de color rojo y la cuarta se de color blanco?

5. En un bolso se encuentran tres monedas; una moneda es corriente, una moneda tiene dos caras y una moneda está cargada de modo que la probabilidad de obtener cara sea 2/3. Se escoge una moneda al azar y se lanza. Hallar la posibilidad p de que salga cara.
6. Se tiene tres cestas con las siguientes especificaciones

   Una cesta A contiene 5 bolas rojas y 7 blancas.
   Una cesta B contiene 3 bolas rojas y 2 blancas.
   Una cesta C contiene 3 bolas rojas y 4 blancas.

   Se selecciona una cesta al azar y se saca una. Si la bola es de color rojo, ¿cuál es la probabilidad de que proceda de la cesta A?
7. La urna A contiene nueve barajas numeradas de 1 a 8, y la urna B contiene cinco barajas numeradas de 1 a 6. Se escoge una urna al azar y se saca una carta. Si el número es impar, hallar la probabilidad de que la carta proceda de la urna A.
8. En una carrera de caballos *a, b,* y *c,* en la primera vuelta sus probabilidades de ganar son, 2/3 y 1/6 respectivamente. Si los caballos corren dos veces, ¿cuál es la probabilidad de que gane c en la primera vuelta?
9. Para la presentación de un proyecto, los estudiantes se escogen al azar, uno al otro, para presentar el proyecto. Calcule la probabilidad p de que hombres y mujeres queden alternados si, (i) la clase consta de 10 hombres y 7 mujeres y (ii) la clase consta de 8 hombres y 13 mujeres.
10. A un jugador le reparten 5 cartas, una tras otra, de una baraja corriente de 52 cartas. ¿Cuál es la probabilidad p de que todas sean espadas?

CAPITULO V

DEMOGRAFÍA

1. EN MORFOLOGÍA DE LA POBLACIÓN

Como se señala al principio del libro, se aborda santamente el uso de las técnicas esbozadas, en el examen de la morfología y dinámica de la población y otros campos de la salud pública.

Cualquier conglomerado humano, presenta en diferentes momentos características muy definidas que constituyen morfología típica. Esta puede ser observada transversalmente por medio de las investigaciones de carácter censal, que permiten estudiar la magnitud de la población en un momento o en una sucesión de estantes, su distribución geográfica composición según atributos.

En la tabla 25 se presenta la evolución de la población de la república de Panamá por provincias desde 1911 hasta 1970. La aceleración demográfica es notable en el total y en algunas provincias, como por ejemplo Panamá, Colon y Chiriqui. La desigualdad relativa de la población de las provincias con respecto al total según los 7 censos queda plasmado en la distribución porcentual correspondiente en cada uno de estos años. Las últimas columnas muestran las grandes diferencias existentes en la densidad demográfica de las provincias. Estos datos sirven para destacar que a mayor o menor población absoluta no necesariamente corresponden a una mayor o menor densidad demográfica. Este breve y simple examen es posible realizar mediante la observación de datos absolutos y cifras relativas como la densidad (razón) porcentaje.

Tabla 25.
Resultados del censo

Censo	Provincia	Frec. relat. 1911	Frec. relat. 1970	Frec. acum. 1911	Frec. acum. 1970	Diferencia entre censos	Cociente intercensal
1	Panamá	11.9	27.2	11,9%	27,2%	15,3%	2,3
2	Herrera	19.1	15.9	31,0%	43,1%	-3,2%	0,8
3	Chiriquí	14.5	14.4	45,5%	57,5%	-0,1%	1,0
4	Coclé	14.1	12.4	59,6%	69,9%	-1,7%	0,9
5	Los santos	15.7	9.9	75,3%	79,8%	-5,8%	0,6
6	Colón	8.6	9.6	83,9%	89,4%	1,0%	1,1
7	Veraguas	10.9	7.3	94,0%	96,7%	-2,8%	0,7
8	Bocas del toro	5.0	2.6	99,0%	99,3%	-2,4%	0,5
9	Darién	1.0	0.7	100,0%	100,0%	-0,3%	0,7
Total		100.0	100.0				

Fuente: Elaboración propia.

Sin embargo, puede realizarse una medición algo más rigurosa (ver tabla 25). Se estudia sólo los totales de la república de Panamá, donde se puede observar el porcentaje de cambio. También representa los cocientes de aumentos íntercensal. En este caso los aumentos absolutos son relacionados a la semisuma de la población en momentos sucesivos.

El análisis de la tabla 25 evidencia las frecuencias relativas y acumuladas de la población de cada una de las provincias, así como la diferencia o cocientes intercensal en cada una de las provincias. Cuando dichos períodos son equidistantes, el aumento absoluto permite ver con claridad la evolución en crecimiento.

Los porcentajes de cambio expresan las fracciones de cambio que se puede atribuir a cada 100 personas existentes al principio del período correspondiente.

Los cocientes intercensales toman en cuenta el hecho de que no solo los individuos presentes al comienzo del período contribuyen al crecimiento, sino también a aquellos que se incorporan durante el mismo. Por ello tienen como denominador a la población media de los períodos.

Es importante resaltar que la frecuencia relativa (%) de la población de cada provincia se hace con respecto al total nacional como indicador de la repartición bruta de la población total.

Sin embargo, tanto estas formas de análisis, como las densidades (razones presentadas en la última columna) están afectadas por la superficie de cada unidad política administrativa. El carácter global del concepto de densidad de la arbitrariedad de los límites que puedan escogerse para fines de clasificación. Para obviar estas dificultades pueden utilizarse el indicador definido por la revisión:

$$I_i = \frac{\textit{Densidad de la unidad territorial}}{\textit{Suma de las densidades de las unidades territoriales}} = \frac{D_i}{\sum D_i}$$

El indicador I_i representa la importancia que tendría la población de la unidad *i* si todas las unidades territoriales tuvieran la misma superficie.

Además del indicador señalado, la densidad demográfica y la importancia relativa de la población, existen otros indicadores de distribución territorial de la población como índices de la concentración Gini, Baricentro o centro de gravedad de la población y otras medidas algo menos simples que en beneficios de la brevedad y simplicidad del presente libro no se examinan.

En el análisis de la composición de acuerdo a determinadas características es muy útil la aplicación de las razones, porcentajes y algunas medidas de posición. Para conocer la distribución por censos se puede recurrir a porcentajes que representen la población de un sexo respecto al conjunto de ambos sexos, o la relación de masculinidad (tasa o índice de masculinidad).

Estos índices pueden arrojar luces sobre fenómenos migratorios (por ejemplo con relación la migración diferencial por sexo y edad), y resultan útiles para detectar errores de empadronamiento en algunos grupos (si la diferencia en la proporción de sexo no aparece lógicamente explicable).

Para estudiar la distribución por edad de la población, resulta útil calcular los porcentajes que representan los grupos de la población de cada edad respecto al total de los habitantes. A veces interesa, además, conocer los porcentajes de la población que está por arriba o por debajo de determinada edad, lo cual se logra calculando frecuencias relativas acumuladas. Ambos cálculos son representados en la tabla 25, según los censos de 1940 y 1970. Incluso para resumir la información se calcula las mediana conforme a las formulas definida con anterioridad.

La composición de la población se puede estudiar de acuerdo con varias características demográficas, sociales, económicas, etc. Las cuales indican la aplicación de algunas medidas básicas ya definidas, con relación a variables, como el estado conyugal y la participación económica del hombre y la mujer en Panamá, según se desprende del último censo levantado en 1970.

En el caso del estado conyugal se puede usar porcentajes de cada edad con respecto a estados conyugales específicos, proporciones de cada estado conyugal a relación con grupos de edad, y también medidas de posición como los tres primeros cuarteles incluyendo, desde luego, a la edad mediana para cada estado conyugal. Se puede establecer si algunas características acerca de la composición por estado conyugal, cambio en la función de la edad tendencia de la nupcialidad, viudez, etc.

Estos porcentajes permiten medir la intensidad del empleo, en el grupo de población susceptible. Para su elaboración se requiere solamente la clasificación de la población económicamente activa que este ocupada y desocupada.

En forma análoga puede calcularse para cada edad y sexo los porcentajes que representan la población económicamente activa de cada edad y sexo respecto al grupo correspondiente. A esta relación se le denomina *tasa de participación o tasa de actividad* y señala la intensidad de la participación en el trabajo de población de cada edad. Son indicadores de desarrollo económico y son fundamentales para calcular un instrumento analítico importante como lo es la *Tabla de Vida Activa.*

Estas tasa de actividad, en particular la global (que es la correspondiente a todas las edades) está afectada por influencias de carácter demográfico (distinta composición por edad de la población), y por factores económicos y sociales. Por ello, en su análisis la utilidad de la aplicación de técnicas de tipificación, a la cuales se les dedicará breve párrafos más adelantes con ejemplos referentes justamente a tasas de actividad.

2. EN DINÁMICA DE POBLACIÓN

Como es justamente el campo en que más desarrollo ha tenido las medidas elementales descritas en párrafos anteriores se indicarán a continuación una lista de tales medidas acompañadas de ejemplos de cálculos. Antes sin embargo, convendría señalar aspectos importantes a tener en cuenta del cálculo de tasa brutas.

Toda tasa es una forma especial de razón. Por lo tanto, consta de un numerador y un denominador. En el numerador se escribe el número de ciclos o veces que ha sido contado o confirmado un fenómeno en el curso de cierto ciclo de tiempo ocurrido dentro de límites geográficos dados. Ese recuento establece la "incidencia" del hecho que debe relacionarse a la población en que ocurrió tal hecho (denominador).

Corrientemente la estadística debería permitir el conocimiento continuo a través del tiempo de los fenómenos estudiados, que constituyen el numerador y así lo hace, aunque con distintos grados de precisión.

En cambio, para conocer el denominador, que en general se refiere a la población que se vincula con los fenómenos estudiados sería necesario llevar un registro permanente. Esto requiere una labor sofisticada de una precisión ideal, hasta ahora presente en sociedades avanzadas desde el punto de vista de la organización social y de desarrollo económico y cultural.

El denominador es entonces, frecuentemente, una estimación de la población existente en un momento determinado. Por convención, tales estimaciones son referidas a la mitad de los períodos para los cuales se calcula la tasa, de modo que sea una aproximación a la "población media", que corresponde mejor con la frecuencia del fenómeno que trata el numerador.

Debe haber correspondencia entre los fenómenos y la población, o sea, concordancia entre numerador y denominador. Los datos respectivos deben ser definidos, clasificados y tabulados de igual manera, si hay divergencia en estos aspectos las tasas carecen de valor. Las concordancias deben cumplirse en tres aspectos:

a) La naturaleza del hecho

La población tiene que ser compatible con el hecho que ha emanado de ella. Por ejemplo, al calcular una tasa que mida el riesgo de una mujer de quedar embarazada, no debe incluirse en la base a mujeres menores de cinco años, ó de modo más general, a mujeres cuya edades están fuera del período representativo.

b) La zona geográfica a que corresponda el hecho

Los datos del numerador y del denominador deben corresponder a la misma zona geográfica.

c) El tiempo

Los datos del denominador deben abarcar el mismo período al que se refiere el numerador. Considerando la posibilidad de varias alternativas: que la población base sea la observada o calculada al comienzo del período; que sea la del final del período; que se refiera a la mitad del período; que sea una media aritmética de las poblaciones observadas durante el período de referencia.

Parece claro que el uso de una estimación de la población al comienzo del período tiende a exagerar el valor de la tasa, ya que esa población es la que origina el primer hecho que la hace variar. Por lo tanto, esa población inicial, es más pequeña que la que realmente corresponde a todo el período considerado. Por ser más pequeña y por tanto un denominador menor produce una tasa mayor o sobrestimada. Una estimación al final del período produce el efecto inverso. La mejor estimación es aquella que más se aproxima a la población media expuesta al riesgo durante el período de observación. Para muchos fines prácticos la población estimada a mitad del período es la mejor estimación. Con el objeto de hacer más fácil la expresión de una tasa es común multiplicar el cociente por una potencia de 10, que a menudo es el número 1.000.

Medidas de la dinámica poblacional

$$\frac{\textit{TASAS}}{\text{(brutas, específicas)}} = \frac{(*)\text{incidencia (frecuencia) de un hecho} \quad *10^{n}}{(*)\text{población que origina los hechos (estimados a mitad de periodo)}}$$

- **Naturaleza**
- **Zona geográfica**
- **Tiempo**

(*) Concordancia

Dentro de este tipo de tasas se puede citar las medidas de fecundidad, mortalidad, nupcialidad y divorcios. Y estas pueden ser "brutas" y "especificas", según tomen en cuenta toda la población o sub-conjuntos de ésta.

3. MEDIDAS DE FECUNDIDAD

Tasa bruta de Natalidad (TBN)

Es la más elemental y expresa la relación entre el número de nacimientos vivos ocurridos durante un período de tiempo y la población media del mismo período. Se expresa en "*por mil habitantes*". De modo general es útil para observar tendencias y su extenso uso es básicamente debido a su mayor disponibilidad.

Fuera de las limitaciones "*técnicas*" que emanan de la calidad de las estadísticas de nacimientos en cuanto se refieren a errores que afectan su integridad, y por lo tanto, el nivel de la tasa, su principal desventaja es que se ve afectada por la estructura de la población femenina según la edad.

Tasa de Fecundidad General

La tasa bruta de natalidad relaciona los nacimientos vivos de la población no expuesta al riesgo. Una forma de corregir ese aspecto es utilizando como denominador exclusivamente a la población femenina en edad fértil. Esa relación multiplicada por mil es la *tasa de fecundidad general*.

Es obvio que podría también calcularse una tasa de esa índole para población masculina, pero la corriente en demografía, es calcular medidas de fecundidad femenina. La fórmula del cálculo de esta tasa (TFG) es:

$$TFG = \frac{Nacimientos\ \ vivos}{Población\ \ femenina\ \ de\ \ 15\ a\ 49\ años}$$

Medidas de Fecundidad

TBN = TASA BRUTA DE NATALIDAD
(Efecto estructura)

TFE = TASA DE FECUNDIDAD GENERAL
(Grupo expuesto a riesgo)

nfx = TASA DE FECUNDIDAD POR EDAD
(Suma de las tasas por edad)

TFG = TASA GLOBAL DE FECUNDIDAD
(Suma de las tasas por edad)

TBR = TASA BRUTA DE REPRODUCCIÓN ó R´
(Índice sintético, experiencia repetida)

Se podrían introducir otros refinamientos como el de limitar a la población femenina casada, por ejemplo, y así calcular una tasa más específica. Sin embargo, esta tasa es menos útil desde el punto de vista demográfico, en especial por la dificultad en establecer comparaciones.

Tasa de Fecundidad por Edad

Estas tasas son las más útiles para el análisis demográfico, pues permiten considerar a la importante variable edad y por medio de ella la tendencia histórica. Su fórmula es la siguiente:

$$nfx = \frac{nBx}{n\overline{N}_{FX}} * 1000$$

Donde

- nfx = Tasa de fecundidad de mujeres de edad x a x+n
- nBx = Nacimientos vivos tenidos por mujeres de edad x a x+n durante un período dado.
- nN_{FX} = Población media femenina de edades x a x+n

Tasa Global de Fecundidad

La suma de las tasas específicas de fecundidad por edad constituye la tasa global de fecundidad (TGF).

Si las tasas de fecundidad corresponden a edades simples desde 15 hasta 49, basta sumar las 35 tasas para obtener la TGF. Corrientemente las tasas específicas no se calculan para edades simples, si no para ciertos grupos. En esos casos la TGF se obtiene sumando las tasas multiplicadas por la amplitud del grupo correspondiente. Así, todos los grupos fueron quinquenales, la TGF será igual a $5\sum$ n f x.

Esta tasa expresa al total de hijos que tendrá una mujer durante toda su vida reproductiva si estuviera expuesta a los riesgos de fecundidad definidos por las tasas específicas.

Tasa Bruta de Reproducción

Tasa específica por edad que permite arribar a un índice sintético de fecundidad de gran interés en demografía, que es la tasa bruta de reproducción.

Esta medida expresa el número de hijas que daría a luz una recién nacida que llegara a experimentar durante el curso de toda su vida reproductiva los riesgos de fecundidad que corresponden a las estructura de las tasa específicas de fecundidad. Su fórmula es:

$$R' = K \sum nfx$$

Siendo K una constante 0.4878 que resulta del supuesto de que por cada 100 nacimientos vivos femeninos ocurren 105 nacimientos masculinos $\left(\frac{100}{205}\right)$

Esta tasa está exenta de la influencia de la distinta composición por edad de las mujeres en el período reproductivo.

Tabla 26.
Distribución de la tasa de fecundidad

Mujeres por edad	n f x	
	1968	**1960**
15-19	148.0	152.9
20-24	317.4	312.1
25-29	278.2	288.9
30-34	201.3	204.4
35-39	131.9	134.1
40-44	45.6	54.3
45-49	9.6	9.3
	Σ = 1132.0	1156.0

Fuente: Elaboración propia.

Tabla 27.
Cálculo de las medidas

Medidas	Panamá	
	1968	**1960**
b o TBN	40.1	42.1
TFG	183.5	186.5
TFG	$5\frac{1132}{1000} = \underline{5.66}$	$5\frac{1156}{1000} = \underline{5.78}$
R´ O TBR	5066 * 0.4878 = 2.76	5.78* 0.4878 = 2.72

Fuente: Elaboración propia.

Tasa neta de reproducción

Esta medida incorpora al análisis la variable mortalidad. Los valores de ella indicarían la cantidad de hijas que tendría una recién nacida que se expusiera durante toda su vida reproductiva a los riesgos de fecundidad, y a los de mortalidad emanados de las probabilidades de muerte de una tabla de mortalidad.

Su fórmula es la siguiente:

$$R = K\ldots\ nfx\ np\ x$$

Siendo *np x* la probabilidad que tiene mujer de edad *x* de sobrevivir a la edad x + n.

Las medidas Sintéticas

Estas tres últimas tasas, conocidas como tasas resumidas de fecundidad y en esencial la tasa bruta de reproducción, suelen ser utilizadas como indicadores de desarrollo económico social, ya que su valor divide eficientemente grupos de países según su estadio de evolución económica.

Así, países cuya tasa bruta de reproducción es inferior a 2.0 puede decirse que pertenecen al grupo de "más desarrollados" y los que están por encima de ese valor a países "menos desarrollado". La asociación de indicadores de desarrollo a tales valores podría establecer la estrecha correlación existente en la gran mayoría de los casos.

Razones de Natalidad

A veces resulta útil para ciertos analisis conocer los valores de relaciones existentes entre nacimientos con cierta características y el total de nacimientos vivos. Estas relaciones se conocen con el nombre genérico de "*natalidad proporcional*" o razones de natalidad. De modo general puede escribirse:

$$\text{Razón de natalidad} = \frac{\text{Nac. vivos de característica}}{\text{Total de Nac. vivos}} * 100$$

En este caso la variable podría ser el peso al nacer, el nacimiento, duración del matrimonio de los padres, etc.

4. MEDIDAS DE MORTALIDAD (*)

Tasa Bruta de Mortalidad

Es probablemente la más usada debido a su facilidad de cálculo y a su valor en estudios de demografía y Salud Pública. En muchos países su valor está limitado por serias deficiencias en el registro civil que afecta a las estadísticas que sirven de base al numerador de la tasa.

Bajo ciertas condiciones y teniendo presente sus limitaciones podrían considerarse a los valores de esta tasa como primeras aproximaciones del estado de salud de una población.

Su fórmula de cálculo es:

$$m = \frac{D}{\overline{N}} * 1000$$

Siendo *m* la tasa bruta de mortalidad, *D* la cantidad de defunciones durante un período de tiempo y *N* la población media del mismo período.

Debe tenerse precaución al comparar tasas brutas de mortalidad, ya que la mortalidad varía con la edad y esta tasa no toma en cuenta las distintas composiciones de la población según la edad.

En realidad esta tasa es un promedio aritmético cuya ponderación es la composición por edad. Así sería posible que poblaciones que tienen grupos de edades avanzadas proporcionalmente mayores tendrán tasas más altas que aquellas poblaciones "jóvenes" donde el peso de los grupos de edades avanzadas es menor. Esta dificultad en la comparación es a menudo obviada mediante técnicas de tipificación o "estandarización", aunque aplicadas a datos distintos de mortalidad, de esta serie.

Tasa de Mortalidad según la edad

Su especificación permite hacer comparaciones entre poblaciones porque elimina posibles diferencias debidas a distintas composiciones por edad. Sin embargo, es obvio que no elimina influencia de factores asociados al fenómeno de la mortalidad, como la ocupación, por ejemplo.

Estas tasas muy útiles en el análisis de la mortalidad y fundamentales para estructurar las funciones biométricas de tablas de mortalidad y en el cálculo de la tasa neta de reproducción. Su fórmula de cálculo es la siguiente:

$$ {}_n m_x = \frac{nDx}{nNx} * 1000 $$

Siendo ${}_n m_x$ la tasa especifica de mortalidad correspondiente a una población con edades x a x + n.

nDx es el total de defunciones de un grupo de personas de edad x a x + n ocurridas durante un período de tiempo, y *nNx* es la población media en el grupo específico de edad x a x + n.

Aunque parezca innecesario es conveniente destacar que la especificación puede ser con relación a variables distintas de la edad, como la ocupación, sexo, ecuación, religión, etc. No obstante, como es difícil cumplir con el requisito de la concordancia entre numerador y denominador de la tasa, es frecuente en estos casos calcular proporciones de muertes con características respecto al total de defunciones.

Tasa de Mortalidad por Causa

Si bien es de mayor importancia el estudio de la mortalidad por causas, (enfermedades y traumatismos) es difícil de efectuar por la complejidad de los fenómenos patológicos que se hallan relacionados a factores de etiología, localización y otros de interrelación orgánica que hacen casi imposible disponer de una clasificación fija y lógica. Sin embargo, justamente porque es muy vasto el campo de la patología, se requiere clasificar las enfermedades, para medir a través de tasas, condiciones sanitarias, avances tecnológicos y eficiencia del servicio.

Convencionalmente se utiliza la clasificación internacional originada en los trabajos de Farr y D´Espine (1855), y que culminaron con la presentación del sistema de clasificación por Bertillon (1903) al Instituto Internacional de Estadística y posterior implantación en 1893. Esta clasificación es revisada periódicamente para mantenerla al día de los progresos. Estas revisiones, necesarias, pueden interrumpir la continuidad de una serie de tasas por causas individuales y por ello con cierta frecuencia se hace indispensable trabajar con grupo de enfermedades.

La mayoría de las tasas según causa de muerte se calculan sobre la base de la población total, y se expresan "por 100.000 habitantes". Y su fórmula se puede expresar:

$$\text{Tasa de mortalidad por causa} = \frac{\text{Total de defunciones debidas a la causa i}}{\text{Población medida}} * 100000$$

Que expresada simbólicamente es: $m = \frac{Di}{\overline{N}} * 1000$

CONCLUSIONES

El alcance del objetivo general se da de acuerdo al diseño del modelo didáctico para la enseñanza y aprendizaje de la estadística, generado a través de la teoría construida a lo largo de esta obra; un modelo didáctico estratégico que utilizó las estrategias pedagógicas y tecnológicas fundamentadas en la enseñanza y aprendizaje de la estadística, lo que permite dar un aporte de insumos al proceso de enseñanza de la estadísticas descriptiva e inferencial, facilitando la trascendencia de estos temas en los campos del saber y enfocar la realidad desde una perspectiva novedosa.

Se estudiaron los fundamentos y los obstáculos epistemológicos de la estadística, concluyendo que la aplicación de ellos genera confianza en los procesos de enseñanza y aprendizaje, porque estos generan un gran número de acciones encaminadas a la construcción de una sólida comprensión en los estudiantes. En este sentido se puede afirmar que en la enseñanza es necesario abordar e incorporar los teoremas fundamentales, la lógica requerida, el método utilizado, la práctica pedagógica y el propósito para la aplicabilidad y funcionalidad de la estadística.

Al identificar los estilos de aprendizaje de los estudiantes de Matemáticas e Ingeniería de la Universidad del Atlántico, se destaca la importancia de reconocer dichos estilos (activo, reflexivo, teórico y pragmático) y su predominio en los educandos, permitiendo seleccionar y aplicar estrategias encaminadas a generar interactividad entre estudiantes y docentes entorno a la estadística.

Los aportes de los autores de este proyecto se ven reflejado desde el primer hasta el último capítulo de este libro. Estos son:

- El capítulo I se diseñó para que el estudiante o investigador tenga un bosquejo general de lo que es estadística, sus ramas, clasificación, definición, la importancia y aplicación en otras ciencias.
- En el capítulo II se explica sobre la recolección, organización, tratamiento y representación de los datos y variables.
- El capítulo III comprende las medidas estadísticas, en especial las tendencias centrales y de variabilidad.
- En el capítulo IV se enseña a utilizar técnicas de conteo, con el objeto de determinar el tamaño de una población o muestra, la probabilidad desde el punto de vista moderno y lo que implica el uso de teoremas y axiomas fundamentales.
- Y el capítulo V hace un estudio detallado sobre demografía, en donde se analiza la morfología y dinámica de la población, medidas de población, medidas de fecundidad y mortalidad.

Se concluye que con este libro combinado con los estilos de enseñanza son muy influyentes en el proceso de aprendizaje de los estudiantes; la aplicación de un estilo de enseñanza tradicional tiene muchas limitaciones para el ritmo de aprendizaje de los estudiantes actuales y para subsanar esto se diseñó esta obra como una herramienta para fortalecer los procesos de enseñanza y aprendizaje. Es importante resaltar que es necesario cambiar o actualizar las prácticas de enseñanza, se debe asumir un rol de docente actualizado, con una actitud favorable hacia la innovación, la capacitación permanente; no resistirse al cambio, ser solícitos para la adecuación de estructuras y mejoramiento de los procesos de enseñanza, porque los docentes son los primeros en ser llamados a capacitarse para generar cambios y transformaciones desde el aula de clase, en especial en la enseñanza de las matemáticas.

RECOMENDACIONES

La universidad debe fortalecer y fomentar en los docentes del área de estadística la actualización y profundización de sus prácticas pedagógicas o estilos de enseñanza, con el objeto de concentrar en sus clases o prácticas educativas el uso de las nuevas estrategias de enseñanza afines al dinamismos y participación de los estudiantes para así propiciar un puente que facilite el diálogo entre los actores del proceso educativo y el conocimiento de la estadística.

Es importante resaltar que los descubrimientos o herramientas innovadoras en los procesos de enseñanza y aprendizaje deben ser acogidos por la academia. Es por esto que la Universidad, se debe apropiar de este contexto para ofrecerlo a la comunidad en general, replicar las experiencias de aprendizaje que hacen activa la reflexión permanente de los docentes y estudiantes que permite la construcción de su aprendizaje y renovar las estructuras, fortaleciendo permanentemente la adquisición de herramientas didácticas.

La Universidad debe fomentar una cultura que propenda por la innovación educativa, para que todos los miembros de la comunidad académica se identifiquen con este motor académico, generando así el sentido de pertinencia con la institución, preponderando así la buena comunicación para apalancar la continuidad de los procesos que apoyan a la excelencia académica.

REFERENCIAS

Bertillon, J. (1903). *Nomenclatures de Maladies.* (Ed.) Commission Internationatle des nomenclatures Nosologiques. Montevrain, p. 95.

Canavos, G. (1999). *Probabilidad y Estadística. Aplicaciones y Métodos.* México, D.F.: McGraw Hill.

Estrella, A. (2011). Historia de la Estadística. *Revista Actualidad Estadística.* (1). 6-12.

Farr, W. y D´Espine, M. (1853). Nomenclatura uniforme de causas de muerte. En J. Bertillon, Congreso Internacional de Estadística. París

Freund, J., Miller, I. y Miller, M. (2000). *Estadística Matemática con aplicaciones.* 6 Ed. México, D.F.: Pearson-Prentice Hall.

Hald, A. (2003). *A History of Probability and Statistics and Their Applications before 1750.* Hoboken, NJ: Wiley.

Hernández, J. (2013). *Breve historia de la estadística.* [Tesis]. Universidad Nacional. Bogotá, D.C.

Martínez, R. (2009). *Estadística básica para topografía. Colección manuales uex–66.* Cáceres: Universidad de Extremadura.

Navidi, W. (2006). *Estadística para ingenieros y científicos*. McGraw Hill.

Walpole, R., Myers, R. y Myers, S. (1999). *Probabilidad y Estadística para ingenieros*. México, D.F.: Prentice Hall Hispanoamericana. México.

Alejandro Urieles GUERRERO.
Licenciado en Matemáticas y Física, Especialista en Matemáticas Avanzada, Magister en Matemáticas y Doctor en Matemáticas. Docente e investigador Universidad del Atlántico en la línea de análisis. Autor de varias publicaciones en Teoría de Aproximación.

William RAMIREZ QUIROGA.
Licenciado en Educación Básicas con énfasis en Matemática y Magister Ciencias Matemáticas. Docente e investigador Universidad de la Costa CUC en la línea de análisis. Autor de varias publicaciones en Funciones Especiales.

Julio Cesar ROMERO PABON.
Licenciado en Matemáticas y Física, Especialista en Docencia y Especialista en Administración Universitaria, Magister en Matemáticas Aplicada y Doctor en Ciencias de la Educación. Docente e investigador de la Universidad del Atlántico.